# Optimizing the specificity of nucleic acid hybridization

## David Yu Zhang, Sherry Xi Chen, and Peng Yin

# Text S1. Concentration-adjusted standard free energy $\Delta G'$ and hybridization yield

We first consider a standard $X + C \rightarrow XC$ hybridization reaction. Let's assume the initial concentrations of $X$ and $C$ to be $[X]_0 = ac$ and $[C]_0 = c$. For simplicity, let's define $x \equiv [XC]_\infty$, the equilibrium concentration of $XC$. Without loss of generality, we assume that $a \geq 1$, and the hybridization yield $\chi = \frac{[XC]_\infty}{[XC]_\infty + [C]_\infty} = \frac{x}{c}$. (If $a < 1$, then $\chi = \frac{[XC]_\infty}{[XC]_\infty + [X]_\infty} = \frac{x}{ac}$, but the mathematics work out similarly.) The value of $x$ must satisfy the equilibrium constant equation.

$$
\begin{aligned}
K_{eq} &= \frac{[XC]_\infty}{[X]_\infty [C]_\infty} \\
&= \frac{x}{(ac - x)(c - x)} \\
K_{eq}(ac - x)(c - x) &= x \\
x &= \frac{(a+1)cK_{eq} + 1}{2K_{eq}} - \frac{\sqrt{(a-1)^2(cK_{eq})^2 + 2(a+1)cK_{eq} + 1}}{2K_{eq}} \\
\chi = \frac{x}{c} &= \frac{(a+1)cK_{eq} + 1}{2cK_{eq}} - \frac{\sqrt{(a-1)^2(cK_{eq})^2 + 2(a+1)cK_{eq} + 1}}{2cK_{eq}}
\end{aligned}
$$

The value of the hybridization yield for various values of $a$ are plotted in Fig. S1a, and for various operational concentrations $c$ in Fig. S1b. As can be seen, the hybridization yield depends only on the concentration adjusted equilibrium constant $K'_{eq} = cK_{eq}$.

$$
\chi = \frac{(a+1)K'_{eq} + 1}{2K'_{eq}} - \frac{\sqrt{(a-1)^2(K'_{eq})^2 + 2(a+1)K'_{eq} + 1}}{2K'_{eq}}
$$

The hybridization yield is plotted against the concentration adjusted standard free energy $\Delta G' = -RT\ln(K'_{eq})$ in Fig. S1c.
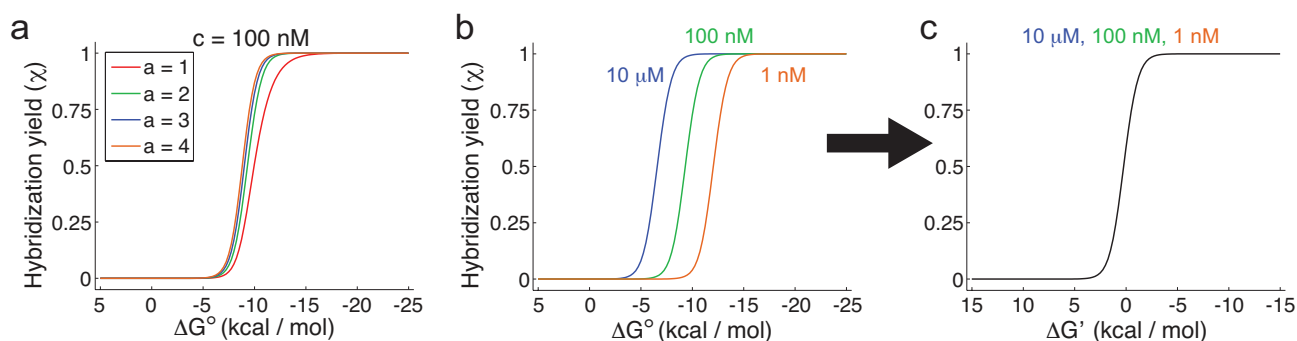
FIG. S1: Hybridization yield analysis of a standard $X + C \rightarrow XC$ hybridization reaction. **(a)** The hybridization yield is sigmoidal, and the exact shape depends on the stoichiometric ratio of $X$ and $C$. **(b)** At different operational concentrations (shown are 1:1 stoichiometry plots), the hybridization yield varies. **(c)** The dimensionless concentration-adjusted standard free energy ($\Delta G'$) effectively reflects the hybridization yields of the system.

**Extension to arbitrary hybridization probes**

Similar analysis can be done for an arbitrary hybridization probe reaction of the form:
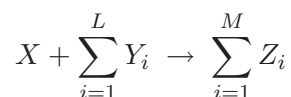
$$X + \sum_{i=1}^{L} Y_i \;\rightarrow\; \sum_{i=1}^{M} Z_i$$

We will assume $[X]_0 = c$, $[Y_i]_0 = a_i c$, and $[Z_i]_0 = b_i c$ for all $i$, with $a_i \geq 1$ and $b_i \geq 0$. At equilibrium, $[Z_i] = b_i c + x$, and the hybridization yield is still defined as $\chi = \frac{x}{c}$.

$$
\begin{aligned}
K_{eq} &= \frac{\prod_{i=1}^{M}[Z_i]}{[X]\prod_{i=1}^{L}[Y_i]} \\
&= \frac{\prod_{i=1}^{M}(b_i c + x)}{(c-x)\prod_{i=1}^{L}(a_i c - x)} \\
c^{L+1-M}K_{eq} &= \frac{\prod_{i=1}^{M}(b_i + \chi)}{(1-\chi)\prod_{i=1}^{L}(a_i - \chi)}
\end{aligned}
$$

The change in the number of species in the reaction is $\Delta n = M - (L+1)$, so the left-hand side can be rewritten as:

$$K'_{eq} \equiv c^{-\Delta n}K_{eq} \;=\; \frac{\prod_{i=1}^{M}(b_i + \chi)}{(1-\chi)\prod_{i=1}^{L}(a_i - \chi)}$$

Solving this equation will yield an expression for $\chi$ that can be written solely in terms of $K'_{eq}$ and constants $L$, $M$, $a_i$ and $b_i$. Consequently, the concentration adjusted equilibrium constant $K'_{eq}$ is an effective sole metric of hybridization yield.

# Text S2. Optimal discrimination factor $Q_{\mathbf{max}}$ predicted by $\Delta\Delta G^\circ$

Given that a correct target and a spurious target hybridize to a complement with standard free energy differing by $\Delta\Delta G^\circ$ (see Fig. 1c, Tables S1-S5), we can calculate the corresponding change in equilibrium constant as $z \equiv Q_{\max} \equiv e^{\Delta\Delta G^\circ/RT}$. Here, we show that this change in equilibrium constant places an upper bound on the discrimination factor achievable. For convenience, we define $K \equiv K_{eq}$ to be the equilibrium constant.

$$K = \frac{x}{(ac-x)(c-x)} \tag{1}$$

$$(ac-x)(c-x)K = x \tag{2}$$

$$-(c-x)Kdx - (ac-x)Kdx + (ac-x)(c-x)dK = dx \tag{3}$$

$$dK = \frac{(a+1)cK - 2xK + 1}{(ac-x)(c-x)}dx \tag{4}$$

$$x = c\chi$$

$$dx = cd\chi$$

Recall that $\chi$ is a function of $K$; here $\chi(K)$ denotes the hybridization yield of a reaction with a particular $K_{eq}$ value. We want to show that:

$$Q \equiv \frac{\chi_X}{\chi_S} = \frac{\chi(zK_{eq,S})}{\chi(K_{eq,S})} < z$$

$$\frac{\frac{\chi(zK_{eq,S})}{zK_{eq,S}}}{\frac{\chi(K_{eq,S})}{K_{eq,S}}} < 1$$

Because $z > 1$, it suffices to show that:

$$\frac{d}{dK}\left(\frac{\chi(K)}{K}\right) < 0$$

$$Kd\chi - \chi dK < 0$$

$$K\frac{dx}{c} - \frac{x}{c}dK < 0$$

$$Kdx < xdK$$

Substituting equations (1) and (4) for $K$ and $dK$,

$$\frac{x}{(ac - x)(c - x)}dx \;<\; x\frac{(a + 1)cK - 2xK + 1}{(ac - x)(c - x)}dx$$

$$1 \;<\; (a + 1)cK - 2xK + 1$$

$$2x \;<\; (a + 1)c$$

Recall that $a \geq 1$, and $c \geq x$; consequently, the above inequality must be true. Thus, $z \equiv Q_{max}$ sets a limit on $Q$ (with $Q < z$ for any $K$).

**$Q_{max}$ is a tight upper bound.** Specifically, we prove that $Q \to Q_{max}$ as $\Delta G' \to +\infty$ for the $X + C \to XC$ and the $X + PC \to XC + P$ reactions. First, for $X + C \to XC$:

$$\begin{aligned}
\lim_{K \to 0} Q &= \lim_{K \to 0} \left(\frac{\chi_X}{\chi_S}\right) \\
&= \lim_{K \to 0} \left(\frac{(a + 1)czK + 1 - \sqrt{(a - 1)^2(czK)^2 + 2(a + 1)czK + 1}}{z((a + 1)cK + 1 - \sqrt{(a - 1)^2(cK)^2 + 2(a + 1)cK + 1})}\right) \\
&= \frac{(a + 1)czK + 1 - (1 + (a + 1)czK - 2ac^2z^2K^2)}{z((a + 1)cK + 1 - (1 + (a + 1)cK - 2ac^2K^2))} \\
&= \frac{2ac^2z^2K^2}{z(2ac^2K^2)} = z
\end{aligned}$$

Consequently, $z = e^{\Delta\Delta G^{\circ}/RT}$ sets the upper bound for discrimination factor for a standard hybridization reaction, and this upper bound is approached as $K_{eq}$ approaches 0, and $\Delta G'$ approaches $+\infty$.

Next, for $X + PC \to XC + P$, with $[X]_0 = ac$, $[PC]_0 = c$, $[XC] = 0$, and $[P] = bc$, we get:

$$\chi = \frac{(a + 1)K + b}{2(K - 1)} - \frac{\sqrt{((a + 1)K + b)^2 - 4a(K - 1)K}}{2(K - 1)}$$

$$\begin{aligned}
\lim_{K \to 0} Q &= \frac{(a + 1)zK + b - \sqrt{b^2 + (2(a + 1)b + 4a)zK + (a^2 + 6a + 1)z^2K^2}}{(a + 1)zK + b - \sqrt{b^2 + (2(a + 1)b + 4a)K + (a^2 + 6a + 1)K^2}} \\
&= \frac{(a + 1)zK + b - (b + (a + 1)zK + \frac{2azK}{b})}{(a + 1)K + b - (b + (a + 1)K + \frac{2aK}{b})} \\
&= \frac{-\frac{2azK}{b}}{-\frac{2aK}{b}} \\
&= z
\end{aligned}$$

with the exception of $b = 0$, where

$$\begin{aligned}
\lim_{K \to 0} Q &= \frac{(a + 1)zK - \sqrt{4azK + (a^2 + 6a + 1)z^2K^2}}{(a + 1)zK - \sqrt{4aK + (a^2 + 6a + 1)K^2}} \\
&\approx \frac{(a + 1)zK - \sqrt{4azK}}{(a + 1)K - \sqrt{4aK}} \\
&\approx \frac{-\sqrt{4azK}}{-\sqrt{4aK}} \\
&= \sqrt{z}
\end{aligned}$$

**Discrimination factor at $\Delta G' = 0$**

When $\Delta G' = 0$, $K'_{eq,X} = 1$, and $K'_{eq,S} = \frac{K'_{eq,X}}{z} = \frac{1}{z}$. We wish to determine a lower bound for discrimination factor $Q = \frac{\chi_X}{\chi_S}$, so we will compute a lower bound for $\chi_X$ and an upper bound for $\chi_S$.

Hybridization yield of $X$ can be calculated as: $\chi_X = 1 - \frac{\sqrt{a^2+4}-a}{2}$. Calculation of $\frac{d\chi_X}{da}$ shows that $\chi_X$ monotonically increases with $a$, so $\chi_X = 0.38$ is smallest value attainable (at $a = 1$).

$$
\begin{aligned}
K_{eq,S} = \frac{1}{cz} &= \frac{x}{(ac-x)(c-x)} \\
x &= \frac{(ac-x)(c-x)}{cz} \\
&< \frac{ac}{z} \\
\chi_S = \frac{x}{c} &< \frac{a}{z}
\end{aligned}
$$

Consequently, $Q = \frac{\chi_X}{\chi_S} > (\frac{0.38}{a})z$; the discrimination factor at $\Delta G' = 0$ is within a constant factor of the theoretical maximum of $z$. For a 1:1 stoichiometry, $Q > 0.38z$, within a factor of 3 of $z$.

**Discrimination factor at melting temperature**

At the melting temperature, $\chi_X = 0.5$.

$\chi_S < \frac{1}{z}$

$Q = \frac{\chi_X}{\chi_S} = \frac{0.5}{\chi_S} > \frac{z}{2}$

## Text S3. Relative thermodynamics of single-base changes

The standard free energy of a hybridization reaction is increased (made more positive) when a spurious target possesses even a single-base change relative to the correct target. The relative thermodynamics of single-base deletions are shown in Table S1, and that of single-base changes are shown in Tables S2-S5, using values taken from SantaLucia and Hicks [1]. The relative thermodynamics of a single-base insertion is +4.00 kcal/mol regardless of the identity of the inserted base or the identities of the flanking bases. Listed $\Delta\Delta G°$ values are calculated as: $\Delta\Delta G° = \Delta G°(SC) - \Delta G°(XC)$, where $SC$ is the spuriously hybridized complex and $XC$ is the correctly hybridized complex. All values are for 1 M Na$^+$ at 37 °C.

As an example, the 5'-ACG-3' sequence normally pairs with 5'-CGT-3', yielding a standard free energy of $\Delta G° = \Delta G°(^{AC}_{TG}) + \Delta G°(^{CG}_{GC}) = -1.44 + (-2.17) = -3.61$ kcal/mol. A deletion of the middle C results in 5'-A-G-3', which yields standard free energy of $\Delta G° = \Delta G°(^{AG}_{TC}) + \Delta G°(\text{bulge}) = -1.28 + 4.0 = +2.72$ kcal/mol, resulting in $\Delta\Delta G° = 2.72 - (-3.61) = +6.33$ kcal/mol. A change of the middle C to a G results in 5'-AGG-3', which yields standard free energy of $\Delta G° = \Delta G°(^{AG}_{TG}) + \Delta G°(^{GG}_{GC}) = -0.13 + (-0.11) = -0.24$ kcal/mol, resulting in $\Delta\Delta G° = -0.24 - (-3.61) = +3.37$ kcal/mol.

The $\Delta\Delta G°$ value sets the maximum discrimination factor, as described previously.

| Deletion (average +5.41) | A | G | C | T |
|---|---|---|---|---|
| A | AAA  +5.00 | AAG  +5.00 | AAC  +5.00 | AAT  +5.00 |
|   | AGA  +5.58 | AGG  +5.84 | AGC  +6.08 | AGT  +5.84 |
|   | ACA  +5.89 | ACG  +6.33 | ACC  +5.84 | ACT  +5.84 |
|   | ATA  +4.46 | ATG  +5.05 | ATC  +4.74 | ATT  +5.00 |
| G | GAA  +5.00 | GAG  +4.74 | GAC  +4.50 | GAT  +4.74 |
|   | GGA  +5.84 | GGG  +5.84 | GGC  +5.84 | GGT  +5.84 |
|   | GCA  +6.39 | GCG  +6.57 | GCC  +5.84 | GCT  +6.08 |
|   | GTA  +4.72 | GTG  +5.05 | GTC  +4.50 | GTT  +5.00 |
| C | CAA  +5.00 | CAG  +4.56 | CAC  +5.05 | CAT  +5.05 |
|   | CGA  +6.02 | CGG  +5.84 | CGC  +6.57 | CGT  +6.33 |
|   | CCA  +5.84 | CCG  +5.84 | CCC  +5.84 | CCT  +5.84 |
|   | CTA  +4.41 | CTG  +4.56 | CTC  +4.74 | CTT  +5.00 |
| T | TAA  +5.00 | TAG  +4.41 | TAC  +4.72 | TAT  +4.46 |
|   | TGA  +6.17 | TGG  +5.84 | TGC  +6.39 | TGT  +5.89 |
|   | TCA  +6.17 | TCG  +6.02 | TCC  +5.84 | TCT  +5.58 |
|   | TTA  +5.00 | TTG  +5.00 | TTC  +5.00 | TTT  +5.00 |

TABLE S1: Relative thermodynamics due to single-base deletions range from +4.41 to +6.57 kcal/mol, with mean of 5.41 kcal/mol. The middle base shown in red is the original base that is deleted. On average, an A or a T deletion results in +4.83 kcal/mol, while a G or a C deletion results in +5.98 kcal/mol.

| A to G (average +2.30) | A | G | C | T |
|---|---|---|---|---|
| A | AAA  +3.05 | AAG  +2.67 | AAC  +2.56 | AAT  +2.66 |
| G | GAA  +2.72 | GAG  +2.34 | GAC  +2.23 | GAT  +2.33 |
| C | CAA  +2.32 | CAG  +1.94 | CAC  +1.83 | CAT  +1.93 |
| T | TAA  +2.35 | TAG  +1.97 | TAC  +1.86 | TAT  +1.96 |
| A to C (average +3.66) | A | G | C | T |
| A | AAA  +3.39 | AAG  +3.32 | AAC  +4.06 | AAT  +3.25 |
| G | GAA  +3.67 | GAG  +3.60 | GAC  +4.34 | GAT  +3.53 |
| C | CAA  +3.82 | CAG  +3.75 | CAC  +4.49 | CAT  +3.68 |
| T | TAA  +3.30 | TAG  +3.23 | TAC  +3.97 | TAT  +3.16 |
| A to T (average +3.08) | A | G | C | T |
| A | AAA  +3.37 | AAG  +2.85 | AAC  +3.58 | AAT  +3.26 |
| G | GAA  +3.43 | GAG  +2.91 | GAC  +3.64 | GAT  +3.32 |
| C | CAA  +3.01 | CAG  +2.49 | CAC  +3.22 | CAT  +2.90 |
| T | TAA  +2.94 | TAG  +2.42 | TAC  +3.15 | TAT  +2.83 |

TABLE S2: Relative thermodynamics due to single-base changes from A to N range from +1.83 to +4.49 kcal/mol, with mean of +3.01 kcal/mol. The middle base shown in red is the original base that is changed. On average, changing an A to a G imposes a +2.30 kcal/mol penalty, changing an A to a C imposes a +3.66 kcal/mol penalty, and changing an A to a T imposes a +3.08 kcal/mol penalty.

| G to A (average +4.98) | A | G | C | T |
|---|---|---|---|---|
| A | AGA  +4.43 | AGG  +4.79 | AGC  +4.87 | AGT  +4.37 |
| G | GGA  +4.92 | GGG  +5.28 | GGC  +5.36 | GGT  +4.86 |
| C | CGA  +5.19 | CGG  +5.55 | CGC  +5.63 | CGT  +5.13 |
| T | TGA  +4.64 | TGG  +5.00 | TGC  +5.08 | TGT  +4.58 |
| G to C (average +5.33) | A | G | C | T |
| A | AGA  +4.96 | AGG  +5.15 | AGC  +5.64 | AGT  +5.38 |
| G | GGA  +4.98 | GGG  +5.17 | GGC  +5.66 | GGT  +5.40 |
| C | CGA  +5.22 | CGG  +5.41 | CGC  +5.90 | CGT  +5.64 |
| T | TGA  +4.85 | TGG  +5.04 | TGC  +5.53 | TGT  +5.27 |
| G to T (average +4.91) | A | G | C | T |
| A | AGA  +4.64 | AGG  +4.47 | AGC  +4.87 | AGT  +4.09 |
| G | GGA  +5.45 | GGG  +5.28 | GGC  +5.68 | GGT  +4.90 |
| C | CGA  +5.20 | CGG  +5.03 | CGC  +5.43 | CGT  +4.65 |
| T | TGA  +4.83 | TGG  +4.66 | TGC  +5.06 | TGT  +4.28 |

TABLE S3: Relative thermodynamics due to single-base changes from G to N range from +4.09 to +5.90 kcal/mol, with mean of +5.08 kcal/mol. The middle base shown in red is the original base that is changed. On average, changing a G to an A imposes a +4.98 kcal/mol penalty, changing a G to a C imposes a +5.33 kcal/mol penalty, and changing a G to a T imposes a +4.91 kcal/mol penalty.

| C to A (average +3.56) | A | G | C | T |
|---|---|---|---|---|
| A | ACA  +3.77 | ACG  +3.86 | ACC  +2.90 | ACT  +2.88 |
| G | GCA  +4.18 | GCG  +4.27 | GCC  +3.31 | GCT  +3.29 |
| C | CCA  +4.06 | CCG  +4.15 | CCC  +3.19 | CCT  +3.17 |
| T | TCA  +3.91 | TCG  +4.00 | TCC  +3.04 | TCT  +3.02 |
| C to G (average +2.94) | A | G | C | T |
| A | ACA  +3.20 | ACG  +3.37 | ACC  +2.04 | ACT  +2.46 |
| G | GCA  +3.02 | GCG  +3.19 | GCC  +1.86 | GCT  +2.28 |
| C | CCA  +3.62 | CCG  +3.79 | CCC  +2.46 | CCT  +2.88 |
| T | TCA  +3.63 | TCG  +3.80 | TCC  +2.47 | TCT  +2.89 |
| C to T (average +3.45) | A | G | C | T |
| A | ACA  +3.39 | ACG  +3.21 | ACC  +3.43 | ACT  +3.50 |
| G | GCA  +3.53 | GCG  +3.35 | GCC  +3.57 | GCT  +3.64 |
| C | CCA  +3.40 | CCG  +3.22 | CCC  +3.44 | CCT  +3.51 |
| T | TCA  +3.52 | TCG  +3.34 | TCC  +3.56 | TCT  +3.63 |

TABLE S4: Relative thermodynamics due to single-base changes from C to N range from +1.86 to +4.27 kcal/mol, with mean of +3.32 kcal/mol. The middle base shown in red is the original base that is changed. On average, changing a C to an A imposes a +3.56 kcal/mol penalty, changing a C to a G imposes a +2.94 kcal/mol penalty, and changing a C to a T imposes a +3.45 kcal/mol penalty.

| T to A (average +3.18) | A | G | C | T |
|---|---|---|---|---|
| A | ATA +2.76 | ATG +3.37 | ATC +2.96 | ATT +3.10 |
| G | GTA +2.88 | GTG +3.49 | GTC +3.08 | GTT +3.22 |
| C | CTA +2.98 | CTG +3.59 | CTC +3.18 | CTT +3.32 |
| T | TTA +2.96 | TTG +3.57 | TTC +3.16 | TTT +3.30 |

| T to G (average +2.41) | A | G | C | T |
|---|---|---|---|---|
| A | ATA +1.90 | ATG +2.38 | ATC +1.95 | ATT +2.04 |
| G | GTA +1.92 | GTG +2.40 | GTC +1.97 | GTT +2.06 |
| C | CTA +2.39 | CTG +2.87 | CTC +2.44 | CTT +2.53 |
| T | TTA +2.74 | TTG +3.22 | TTC +2.79 | TTT +2.88 |

| T to C (average +3.91) | A | G | C | T |
|---|---|---|---|---|
| A | ATA +3.15 | ATG +3.85 | ATC +3.76 | ATT +3.53 |
| G | GTA +3.41 | GTG +4.11 | GTC +4.02 | GTT +3.79 |
| C | CTA +3.57 | CTG +4.27 | CTC +4.18 | CTT +3.95 |
| T | TTA +3.83 | TTG +4.53 | TTC +4.44 | TTT +4.21 |

TABLE S5: Relative thermodynamics due to single-base changes from T to N range from +1.90 to +4.53 kcal/mol, with mean of +3.17 kcal/mol. The middle base shown in red is the original base that is changed. On average, changing a T to an A imposes a +3.18 kcal/mol penalty, changing a T to a G imposes a +2.41 kcal/mol penalty, and changing a T to a C imposes a +3.91 kcal/mol penalty.
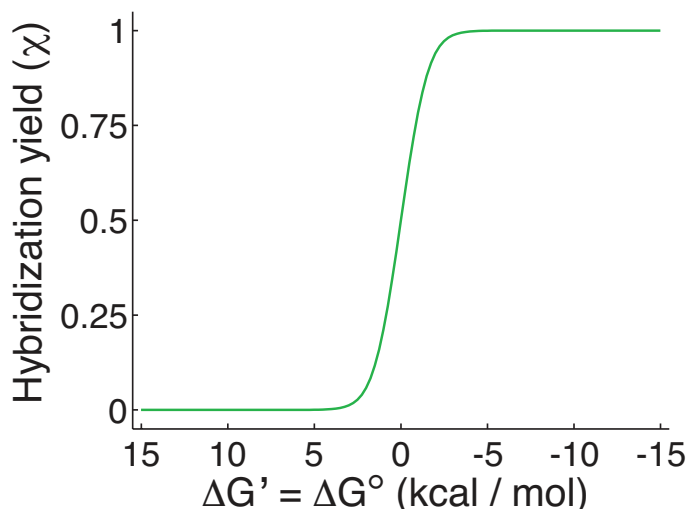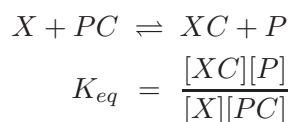
FIG. S2: Hybridization yield for a toehold exchange system with $2\times$ $X$, $1\times$ $P$, and $1\times$ $PC$. For the main paper, $[PC]_0 = 100$ $nM$ was typical.

## Text S4. Calculated hybridization yield for toehold exchange probes

The reaction of a target $X$ with the experimental toehold exchange probe $PC$ can be written as follows:

$$X + PC \; \rightleftharpoons \; XC + P$$
$$K_{eq} \; = \; \frac{[XC][P]}{[X][PC]}$$

For the reactions experimentally tested, we started with $[PC]_0 = c$, $[P]_0 = c$, and $[X]_0 = 2c$. Define $x = [XC]_\infty$ to be the concentration of hybridized product $XC$ at equilibrium. The hybridization yield is then $\chi = \frac{[XC]_\infty}{[XC]_\infty + [PC]_\infty} = \frac{x}{c}$.

$$
\begin{aligned}
K_{eq} \; &= \; \frac{[XC][P]}{[X][PC]} \\[6pt]
&= \; \frac{x(c+x)}{(2c-x)(c-x)} \\[6pt]
x \; &= \; \frac{3cK_{eq} + c}{2K_{eq} - 2} - \frac{c\sqrt{K_{eq}^2 + 14K_{eq} + 1}}{2K_{eq} - 2} \\[6pt]
\chi \; &= \; \frac{x}{c} = \frac{3K_{eq} + 1}{2K_{eq} - 2} - \frac{\sqrt{K_{eq}^2 + 14K_{eq} + 1}}{2K_{eq} - 2}
\end{aligned}
$$

Fig. S2 plots the analytic value of $\chi$ expected for a given reaction $\Delta G' = \Delta G^\circ$.

In the example reaction shown in Fig. 2a, $\Delta G^\circ = 0.51$ kcal/mol at 25 °C, yielding $K_{eq} = 0.423$ and $\chi = 0.343$.
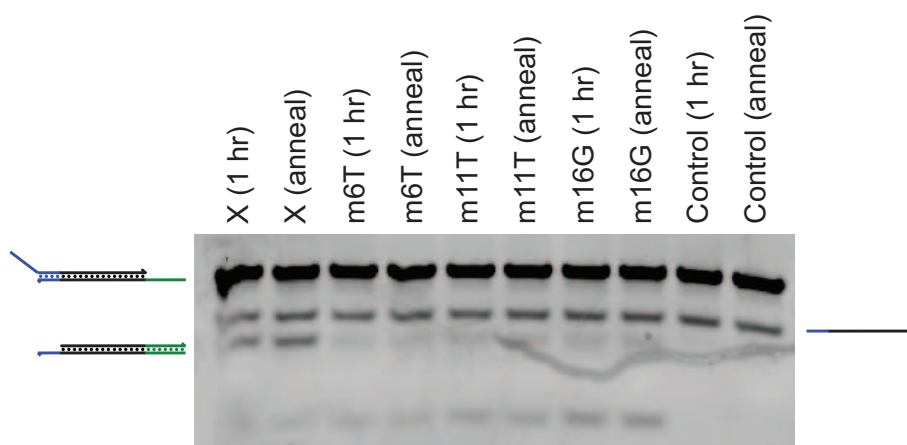
FIG. S3: Comparison of band distribution after 1 hour of reaction and after thermal annealing for the X1-7/6 toehold exchange system. Adjacent lanes show the species distribution after 1 hour of reaction and after thermal annealing. In all cases, band distributions were nearly identical for annealed versus isothermally reacted. This verifies that the reaction between targets and the toehold exchange probe completes in 1 hour. Fig. S7 further shows that even at 1 nM concentration, equilibration is achieved in 20 minutes.
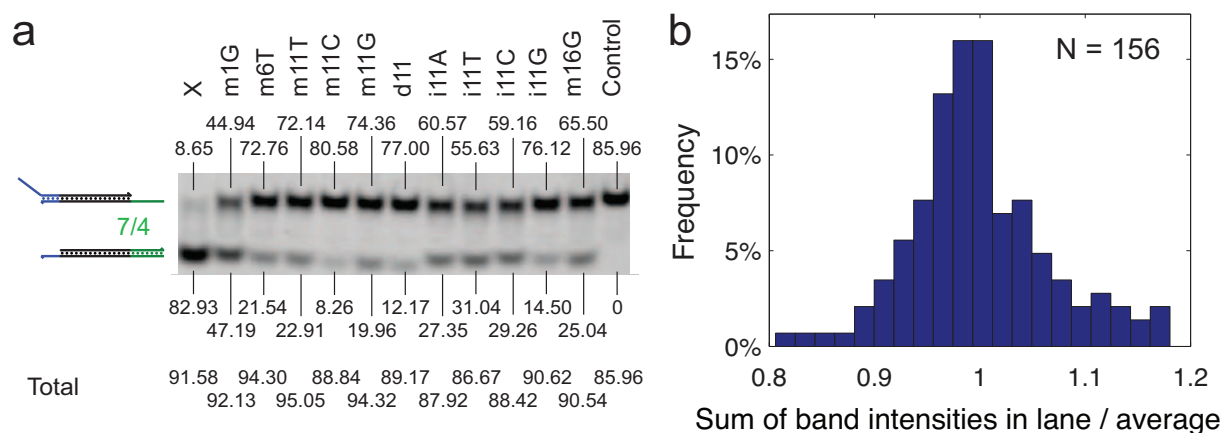


FIG. S4: Method for quantitating hybridization yields from native polyacrylamide gel electrophoresis (PAGE) results. (a) Quantitation of bands in the X1 7/4 system (shown as the second gel in Fig. 3b). The top band is $PC$ and the bottom band is $XC$ or $SC$, and their intensities are marked. Shown at the bottom is the sum of the intensities of the $PC$ and $XC$ bands. (b) Histogram of the sum of intensities of the $PC$ and $XC$ bands in each lane divided by the gel average. For example, the gel average for the gel shown in panel (a) is 90.80, so the data generated by the lanes in panel (a) would be $\frac{91.58}{90.80} = 1.009$, $\frac{92.13}{90.80} = 1.015$, etc. This graphic combines data from the X1, X2, X3, X4, and X5 systems (Fig. 3, Figs. S8-S11). The standard deviation of this distribution is 0.066. Thus, the sum of the $XC$ and $PC$ band intensities are roughly preserved across lanes on the same gel, and hence the band intensities can be directly used to quantitate hybridization yield.

## Text S5. Estimation of strand purity

On the IDT website (http://www.idtdna.com/Catalog/Purification/page1.aspx), the company claims that that HPLC-purified strands will have purity of 85+%, while PAGE-purified strands will have purity of 90+%. This is consistent with the lead author's previous attempts to characterize the purity of oligonucleotides with post-synthesis purification [4], which showed that IDT strands with PAGE purification possessed 92-98% correct-length product. Taking the average of these values (95%), and assuming that the shorter-length oligonucleotides are (n-1)-mers with a uniform distribution of deletions at every position, this results in roughly a 0.2% population of $C$ with deletion at any particular position. For the 0.2% of $C$ that possesses a deletion at position 11, the $d11$ spurious target would be perfectly matched and hybridize with high yield. This results in a minimum hybridization yield of 0.002, making the 7/6 system less sensitive than the 7/5 system.

## Text S6. Calculated hybridization yield for toehold exchange probes

The standard free energies of strands and complexes are needed in order to calculate the standard free energy of reactions, which in turn can be used to generate the equilibrium constant.

For toehold exchange reactions,

$$X + PC \; \rightleftharpoons \; XC + P$$
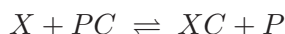
The standard free energy of the reaction is thus calculated as:

$$\Delta G^\circ(\text{reaction}) \; = \; \Delta G^\circ(XC) + \Delta G^\circ(P) - \Delta G^\circ(X) - \Delta G^\circ(PC)$$

The numerical $\Delta G^\circ$ values (partition function) calculated for each strand and complex by NUPACK [2] are listed in Tables S6-S8.

| Name | Protector Sequence | $\Delta G°(P)$ | $\Delta G°(PC)$ | $\Delta G°$ (reaction with Correct) |
|---|---|---|---|---|
| X1-P7/6 | $T_{30}$ TGCATCCACTCATTCAATACC | -0.57 | -32.50 | -0.65 |
| X1-P7/5 | $T_{30}$ GCATCCACTCATTCAATACC | -0.57 | -31.18 | -1.97 |
| X1-P7/4 | $T_{30}$ CATCCACTCATTCAATACC | -0.22 | -28.71 | -4.09 |
| X1-P7/0 | $T_{30}$ CACTCATTCAATACC | -0.19 | -22.62 | -10.15 |
| X2-P7/5 | $T_{30}$ ATGATTGAGGTAGTAGTTTG | -0.40 | -30.14 | -1.73 |
| X2-P7/4 | $T_{30}$ TGATTGAGGTAGTAGTTTG | -0.29 | -28.70 | -3.06 |
| X3-P7/5 | $T_{30}$ AGGATTTAATGCTAATCGTG | -2.64 | -30.49 | -4.25 |
| X3-P7/4 | $T_{30}$ GGATTTAATGCTAATCGTG | -1.36 | -28.92 | -4.54 |
| X4-P7/5 | $T_{30}$ CTCATCACTTGATACAAGCT | -2.61 | -30.28 | -2.48 |
| X4-P7/4 | $T_{30}$ TCATCACTTGATACAAGCT | -2.62 | -28.61 | -4.16 |
| X5-P7/5 | $T_{30}$ CGTTCCAAGAACAGATGTAC | -2.88 | -30.93 | -2.02 |
| X5-P7/4 | $T_{30}$ GTTCCAAGAACAGATGTAC | -3.48 | -28.59 | -4.96 |

TABLE S6: Sequences and energetics of the protector ($P$) and probe ($PC$) species. Column 1 shows the names of the various different protected complements $PC$, with the last number representing the number of bases that must spontaneously dissociate in order for the protector to be released. Column 2 shows the sequence of the protector $P$, with the $T_{30}$ denoting 30 continuous Thymines. The sequence of the complement strand can be derived from the sequence of the protector and that of the Correct target; for example, the sequence of the complement strand for X1-P7/5 would be 'GACGTAGGGTATTGAAT-GAGTGGATGC.' Columns 3 and 4 show the calculated standard free energies of the protector $P$ and the protected complement $PC$, respectively (all energies given in kcal/mol). Column 5 shows the calculated standard free energy of the reaction $X + PC \rightleftharpoons XC + P$, where $X$ is the Correct target of the system (using the $\Delta G°$ values of $X$ and $XC$ given in Tables S7 and S8).

| Name | Sequence | $\Delta G°(X$ or $S)$ | $\Delta G°(XC$ or $SC)$ | $\Delta\Delta G°$ |
|---|---|---|---|---|
| X1-Correct | CACTCATTCAATACCCTACGTC | -0.53 | -33.11 | +0.00 |
| X1-m1G | GACTCATTCAATACCCTACGTC | -0.79 | -32.03 | +1.34 |
| X1-m6T | CACTCTTTCAATACCCTACGTC | -0.47 | -30.12 | +2.93 |
| X1-m11T | CACTCATTCATTACCCTACGTC | -0.48 | -29.74 | +3.32 |
| X1-m11C | CACTCATTCACTACCCTACGTC | -0.49 | -29.56 | +3.51 |
| X1-m11G | CACTCATTCAGTACCCTACGTC | -1.57 | -30.16 | +3.99 |
| X1-d11 | CACTCATTCA-TACCCTACGTC | -0.49 | -28.79 | +4.28 |
| X1-i11A | CACTCATTCAAATACCCTACGTC | -0.62 | -30.18 | +3.02 |
| X1-i11T | CACTCATTCATATACCCTACGTC | -0.58 | -29.60 | +3.56 |
| X1-i11C | CACTCATTCACATACCCTACGTC | -0.57 | -29.52 | +3.63 |
| X1-i11G | CACTCATTCAGATACCCTACGTC | -1.39 | -29.43 | +4.54 |
| X1-m16G | CACTCATTCAATACCGTACGTC | -0.63 | -30.14 | +3.07 |
| X2-Correct | TGAGGTAGTAGTTTGTACAGTT | -1.62 | -33.09 | +0.00 |
| X2-m2C | TCAGGTAGTAGTTTGTACAGTT | -1.66 | -30.05 | +3.08 |
| X2-m7T | TGAGGTTGTAGTTTGTACAGTT | -1.89 | -30.48 | +2.88 |
| X2-m12A | TGAGGTAGTAGATTGTACAGTT | -1.55 | -29.56 | +3.46 |
| X2-m17T | TGAGGTAGTAGTTTGTTCAGTT | -0.65 | -29.66 | +2.46 |
| X2-m17C | TGAGGTAGTAGTTTGTCCAGTT | -0.96 | -28.66 | +3.77 |
| X2-m17G | TGAGGTAGTAGTTTGTGCAGTT | -0.59 | -30.90 | +1.16 |
| X2-d17 | TGAGGTAGTAGTTTGT-CAGTT | -0.69 | -28.58 | +3.58 |
| X2-i17A | TGAGGTAGTAGTTTGTAACAGTT | -1.62 | -29.92 | +3.17 |
| X2-i17T | TGAGGTAGTAGTTTGTTACAGTT | -1.69 | -29.92 | +3.24 |
| X2-i17G | TGAGGTAGTAGTTTGTGACAGTT | -1.25 | -29.64 | +3.08 |
| X2-i17C | TGAGGTAGTAGTTTGTCACAGTT | -1.18 | -29.68 | +2.97 |
| X3-Correct | TTAATGCTAATCGTGATAGGGT | -1.18 | -33.28 | +0.00 |
| X3-m3T | TTTATGCTAATCGTGATAGGGT | -1.25 | -31.17 | +2.18 |
| X3-m8A | TTAATGCAAATCGTGATAGGGT | -1.47 | -29.85 | +3.72 |
| X3-m8G | TTAATGCGAATCGTGATAGGGT | -1.29 | -30.64 | +2.75 |
| X3-m8C | TTAATGCCAATCGTGATAGGGT | -1.73 | -29.33 | +4.50 |
| X3-d8 | TTAATGC-AATCGTGATAGGGT | -1.20 | -29.13 | +4.17 |
| X3-i8A | TTAATGCATAATCGTGATAGGGT | -1.27 | -29.91 | +3.46 |
| X3-i8T | TTAATGCTTAATCGTGATAGGGT | -1.23 | -30.12 | +3.21 |
| X3-i8G | TTAATGCGTAATCGTGATAGGGT | -1.80 | -29.87 | +4.03 |
| X3-i8C | TTAATGCCTAATCGTGATAGGGT | -2.56 | -30.05 | +4.61 |
| X3-m13C | TTAATGCTAATCCTGATAGGGT | -2.55 | -27.39 | +7.26 |
| X3-m18T | TTAATGCTAATCGTGATTGGGT | -1.88 | -30.67 | +3.31 |

TABLE S7: Sequences and energetics of correct targets $(X)$ and spurious targets $(S)$ for the X1, X2, and X3 systems. Column 1 shows the names of the various different targets, and column 2 shows the sequence. Deviations from the correct target are shown in red, and '-' denotes a deletion. Columns 3 and 4 show the calculated standard free energies of the target $T$ and the target-complement product complex $TC$ (all energies given in kcal/mol). Column 5 shows the difference in standard free energies of the reaction with the correct target compared with the listed target.

| Name | Sequence | $\Delta G^\circ(X$ or $S)$ | $\Delta G^\circ(XC$ or $SC)$ | $\Delta\Delta G^\circ$ |
|---|---|---|---|---|
| X4-Correct | CACTTGATACAAGCTTACCATC | -2.51 | -32.66 | +0.00 |
| X4-m4A | CAC**A**TGATACAAGCTTACCATC | -0.89 | -29.02 | +2.02 |
| X4-m9T | CACTTGAT**T**CAAGCTTACCATC | -2.51 | -29.24 | +3.42 |
| X4-m14G | CACTTGATACAAG**G**TTACCATC | -2.97 | -30.26 | +2.86 |
| X4-m19G | CACTTGATACAAGCTTAC**G**ATC | -2.63 | -28.78 | +4.00 |
| X4-m19A | CACTTGATACAAGCTTAC**A**ATC | -2.53 | -28.24 | +4.44 |
| X4-m19T | CACTTGATACAAGCTTAC**T**ATC | -2.54 | -28.95 | +3.74 |
| X4-d19 | CACTTGATACAAGCTTAC-ATC | -2.51 | -27.73 | +4.93 |
| X4-i19C | CACTTGATACAAGCTTAC**C**CATC | -2.53 | -29.71 | +2.97 |
| X4-i19G | CACTTGATACAAGCTTAC**G**CATC | -2.72 | -28.96 | +3.91 |
| X4-i19A | CACTTGATACAAGCTTAC**A**CATC | -2.54 | -28.98 | +3.71 |
| X4-i19T | CACTTGATACAAGCTTAC**T**CATC | -2.55 | -28.88 | +3.82 |
| X5-Correct | CAAGAACAGATGTACCATCACA | -2.75 | -32.82 | +0.00 |
| X5-m3T | CA**T**GAACAGATGTACCATCACA | -2.83 | -30.18 | +2.72 |
| X5-m8T | CAAGAAC**T**GATGTACCATCACA | -3.07 | -30.29 | +2.85 |
| X5-m13A | CAAGAACAGATG**A**ACCATCACA | -2.76 | -29.51 | +3.32 |
| X5-m13C | CAAGAACAGATG**C**ACCATCACA | -2.72 | -28.99 | +3.80 |
| X5-m13G | CAAGAACAGATG**G**ACCATCACA | -2.74 | -30.43 | +2.38 |
| X5-d13 | CAAGAACAGATG-ACCATCACA | -2.12 | -28.38 | +3.81 |
| X5-i13A | CAAGAACAGATG**A**TACCATCACA | -3.30 | -29.36 | +4.01 |
| X5-i13T | CAAGAACAGATG**T**TACCATCACA | -3.16 | -29.65 | +3.58 |
| X5-i13C | CAAGAACAGATG**C**TACCATCACA | -3.11 | -29.24 | +3.94 |
| X5-i13G | CAAGAACAGATG**G**TACCATCACA | -3.63 | -29.66 | +4.04 |
| X5-m18A | CAAGAACAGATGTACCA**A**CACA | -1.64 | -29.57 | +2.14 |

TABLE S8: Sequences and energetics of correct targets ($X$) and spurious targets ($S$) for the X4 and X5 systems.
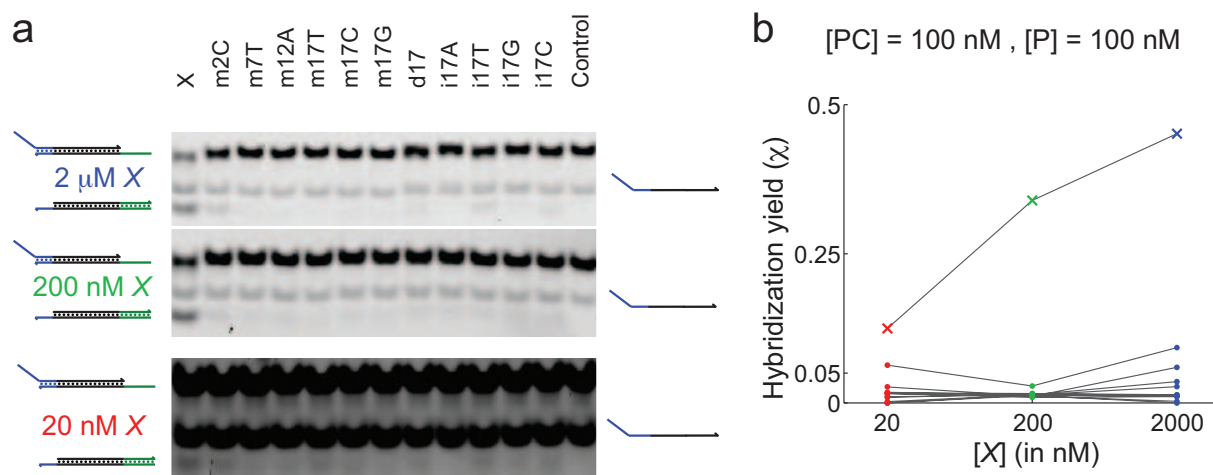
FIG. S5: Performance of the toehold exchange probes with varying concentrations of targets. **(a)** Native PAGE gel results. **(b)** Plot of hybridization yield as a function of target concentrations. The hybridization yield for 20 nM was calculated as $\chi = \frac{5\{XC\}}{\{XC\}+\{PC\}}$ (or its equivalent with $SC$), due to the fact that $X$ or $S$ were the limiting reagents. Due to the low intensities of the $SC$ bands in the 20 nM gel, hybridization yields of spurious targets are less quantitatively reliable than that of other data points.
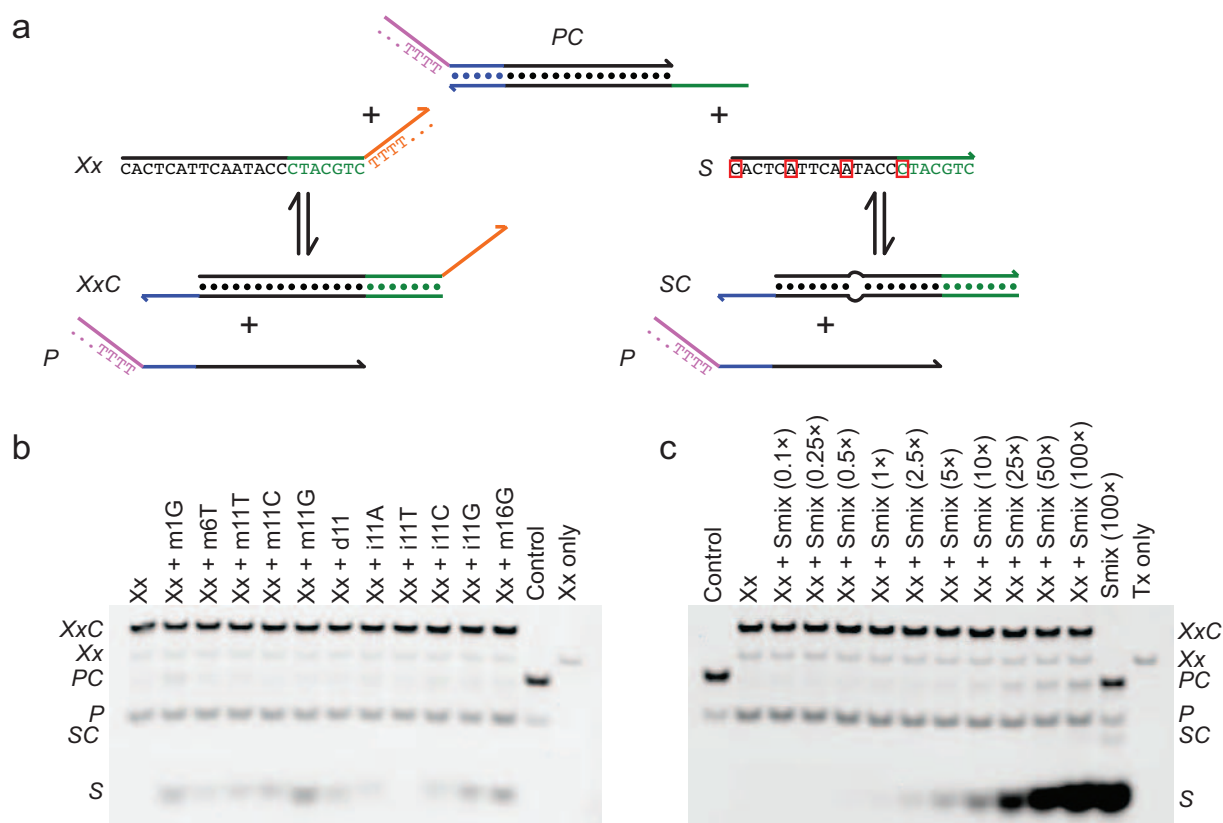
FIG. S6: Effects of heterogeneous samples of correct and spurious targets on hybridization yield. **(a)** Schematic of the experimental design. In order to distinguish $XC$ from $PC$ and $SC$, we appended a 60 nt poly-T tail to the 3' end of the correct target $X$; this modified target is referred to as $Xx$. Recall that the 5' end of protector $P$ has a 30 nt poly-T tail. The probe used has toehold lengths of 7/5. **(b)** 1:1 mixtures of correct and spurious target. Each of lanes 2 through 12 show the reaction between a mixture of the extended target $Xx$ and one spurious target with the toehold exchange probe. Lane 13 ("Control") shows the toehold exchange probe in the absence of any targets, and lane 14 ("Xx only") shows the position of the extended target $Xx$ for reference. The hybridization yield of the extended target $Xx$ was observed to be significantly higher than that of $X$ (from Fig. 3b), implying that the poly-T tail thermodynamically encourages the formation of a nearby duplex either by stabilizing the duplex or by destabilizing the single-stranded $Xx$ molecule. This is consistent with the observation that $P$ seems to bind more strongly to the complement $C$ than predicted by NUPACK. **(c)** Effects of varying amounts of spurious target on the hybridization yield of the correct target. "Smix" denotes an equal mixture of all 12 spurious targets tested for this X1 system. $[PC] = 100$ nM and $[Xx] = 200$ nM $\equiv 1\times$. As can be seen, even at 20 $\mu$M (total concentration of all $S$ species), the mixture of spurious targets still did not bind significantly to the probe. Consequently, the hybridization yield of $Xx$ was not significantly affected by the presence of varying amounts of the spurious target mixture.
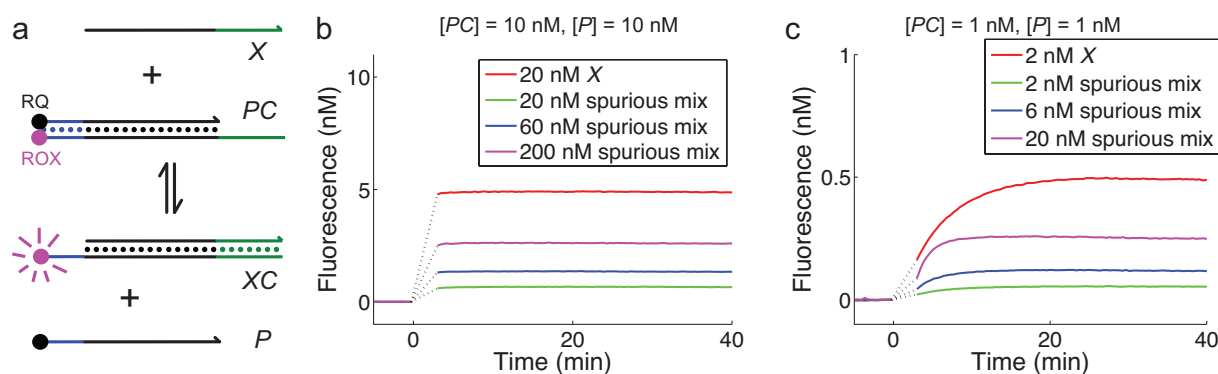
FIG. S7: Fluorescence studies of toehold exchange probes. (a) Schematic of the fluorophore/quencher-labeled probe. ROX denotes the carboxy-X-rhodamine fluorophore, and RQ denotes the Iowa Black Red Quencher (RQ). Upon reaction with a correct or spurious target, the quencher-functionalized protector strand is released, and the fluorescence of the solution is increased. (b) Time-based fluorescence measurement of the operation of the toehold exchange probe at 10 nM concentration. Species $PC$ and $P$ were in the cuvette initially, and correct target or the mixture of spurious targets was introduced at $t \approx 0$. Listed concentration of the Spurious Mix denotes the total concentration of all 11 spurious targets listed. At these concentrations, the toehold exchange reaction had completed in the 5 minutes it took to add target and mix the 4 samples. (c) Time-based fluorescence measurement of the operation of the toehold exchange probe at 1 nM concentration. The rate constant of the reaction is fitted to be $2 \cdot 10^6$ M$^{-1}$ s$^{-1}$; this value is within a factor of 2 of the rate constant of hybridizing two complementary strands of DNA [3]. These results suggest that the correct target will out-compete over 10x excess of spurious targets in hybridizing to the probe.
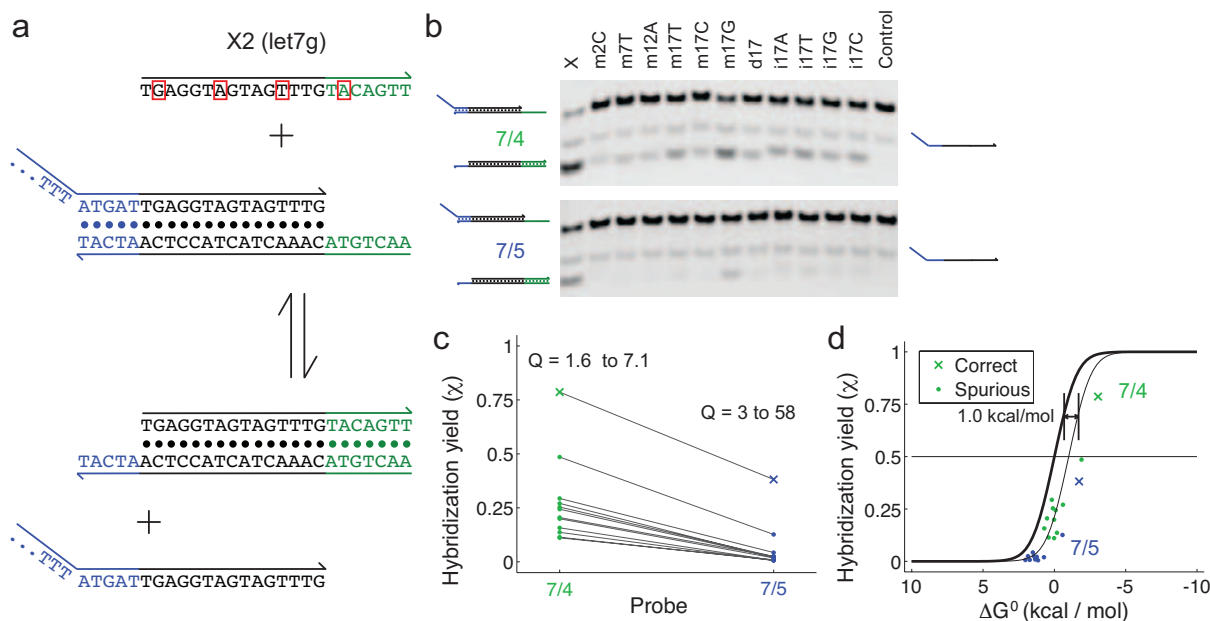
FIG. S8: Experimental results for the X2 system. **(a)** Schematic. Outlined in red are the positions of the one-base changes in various spurious targets. **(b)** Native PAGE results. The protected complement was prepared at a 2:1 ratio of protector to complement, and annealed at 1 $\mu$M concentration of $PC$. The correct or spurious targets were added to achieve final concentration 200 nM of target, 100 nM of protected complement, and 100 nM of free protector. All reactions proceeded at room temperature for 1 hour. The gels show results when the pre-hybridized protector strand must spontaneously dissociate 5 or 4 base pairs to be released; the 7 indicates that the intended target forms 7 new base pairs upon hybridization. The left-most lane shows the reaction between the correct target $X$ and probe $PC$. Middle lanes each shows the reaction between a spurious target $S$ and $PC$; 'm' denotes mismatch, 'd' denotes deletion, and 'i' denotes insertion (e.g. m11C denotes that the A at position 11 was replaced by a C). The right-most lane (labeled 'Control') shows the $PC$ and $P$ solution without the introduction of any target. **(c)** Graphic summary of results from panel (b). Shown in ×'s are the hybridization yields of the correct target $X$; shown in dots are the hybridization yields of the spurious targets $S$. The hybridization yield of the correct target $X$ is calculated as $\chi = \frac{\{XC\}}{\{XC\}+\{PC\}}$, where $\{XC\}$ denotes the band intensity of $XC$. Hybridization yields were similarly calculated for spurious targets $S$. **(d)** Plot of observed hybridization yield vs. yield predicted by reaction thermodynamics. The dark black sigmoidal trace denotes the expected results, and the thin black sigmoidal trace shows the expected results when all reaction standard free energies ($\Delta G^{\circ}$) are adjusted by +1.0 kcal/mol.
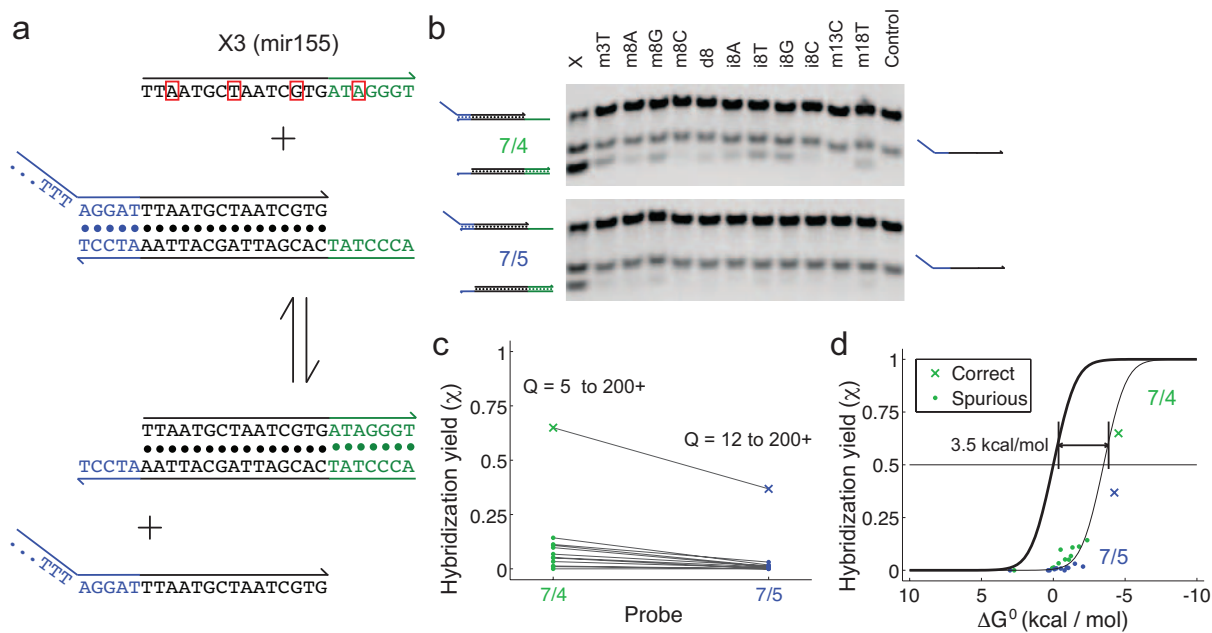
FIG. S9: Experimental results for the X3 system. (a) Schematic. (b) Native PAGE results. (c) Graphic summary of results from panel (b). (d) Plot of observed binding fraction vs. predicted reaction thermodynamics. The thin black sigmoidal trace shows the expected results when all reaction standard free energies ($\Delta G°$) are adjusted by +3.5 kcal/mol.
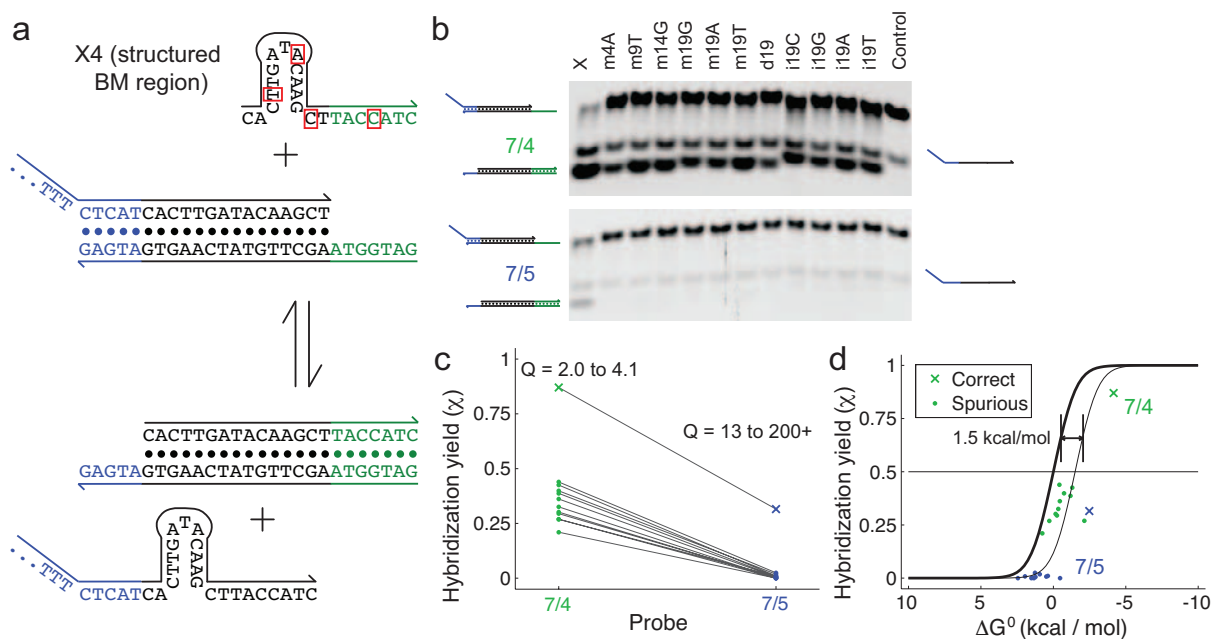


FIG. S10: Experimental results for the X4 system. (a) Schematic. (b) Native PAGE results. (c) Graphic summary of results from panel (b). (d) Plot of observed binding fraction vs. predicted reaction thermodynamics. The thin black sigmoidal trace shows the expected results when all reaction standard free energies ($\Delta G°$) are adjusted by +1.5 kcal/mol.
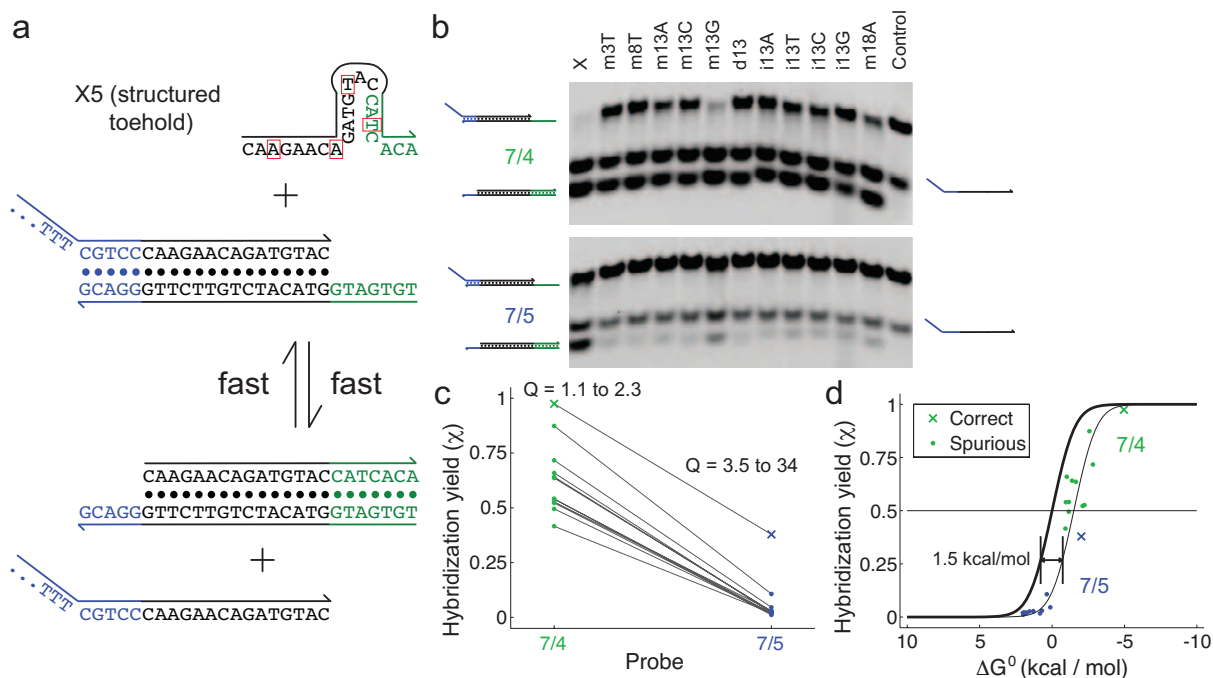
FIG. S11: Experimental results for the X5 system. **(a)** Schematic. **(b)** Native PAGE results. **(c)** Graphic summary of results from panel (b). **(d)** Plot of observed binding fraction vs. predicted reaction thermodynamics. The thin black sigmoidal trace shows the expected results when all reaction standard free energies ($\Delta G^\circ$) are adjusted by +1.5 kcal/mol.
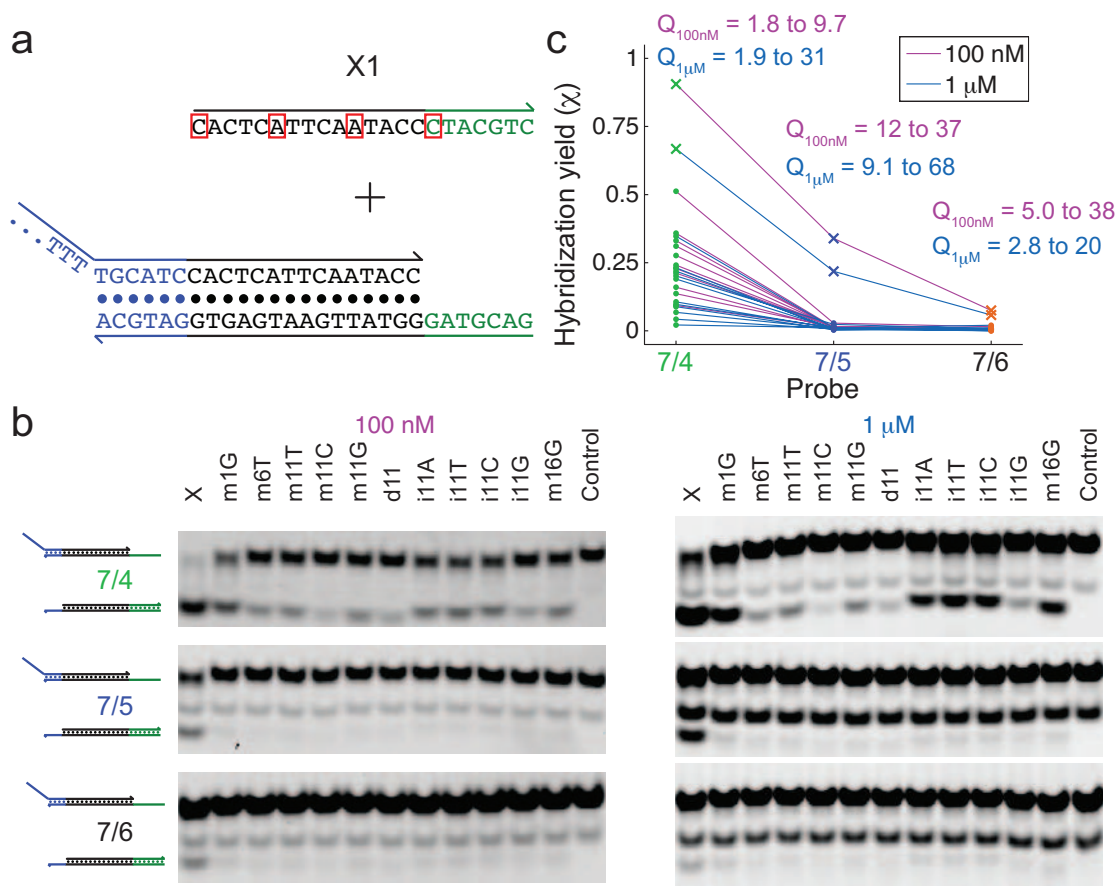
FIG. S12: Concentration dependence of toehold exchange probes. (a) The X1-7/4, X1-7/5, and X1-7/6 systems were tested at two different concentrations. Shown here is the schematic of the X1-7/6 probe. (b) Experimental results. The left gel shows the results when initial concentrations were $[X]=200$ nM and $[P] = [PC] = 100$ nM (data repeated from Fig. 3b), and the right gel shows the results when initial concentrations were $[X] = 2$ $\mu$M and $[P] = [PC] = 1$ $\mu$M. (c) Quantitation of results shown in panel (b).
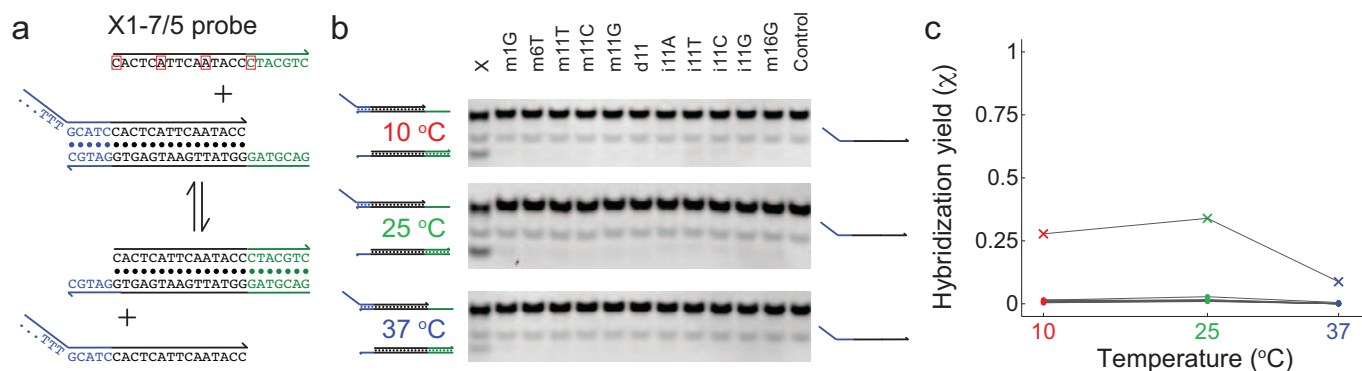


FIG. S13: Temperature dependence of toehold exchange probes. (a) The X1-7/5 system that was tested at two additional temperatures, 10 °C and 37 °C. (b) Experimental results. (c) Quantitation of results shown in panel (b).
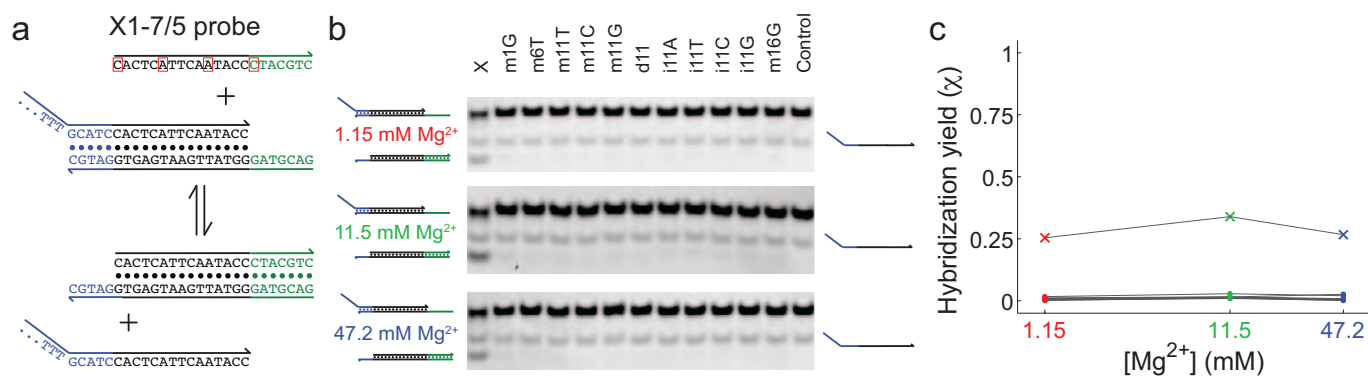
FIG. S14: Salinity dependence of toehold exchange probes. **(a)** The X1-7/5 system that was tested at two additional salt concentrations, 1.15 mM Mg$^{2+}$ and 47.2 mM Mg$^{2+}$. **(b)** Experimental results. **(c)** Quantitation of results shown in panel (b).

[1] SantaLucia, J.; Hicks, D. *Annu. Rev. Biophys. Biomol. Struct.* **2004,** *33*,, 415.

[2] Dirks, R.M., Bois, J.S., Schaeffer, J. M., Winfree, E. & Pierce, N.A. Thermodynamic Analysis of Interacting Nucleic Acid Strands. *SIAM Review* **49,** 65-88 (2007).

[3] Zhang, D.Y. & Winfree, E. Control of DNA Strand Displacement Kinetics Using Toehold Exchange. *J. Am. Chem. Soc.* **131,** 17303-17314 (2009).

[4] Zhang, D.Y. & Winfree, E. Robustness and modularity properties of a non-covalent DNA catalytic reaction. *Nucleic Acids Res.* **38,** 4182-4197 (2010).