Electronic Supplementary Material (ESI) for Analyst.

Supporting Information for

Small molecule-protein interactions in branch migration thermodynamics: modelling and application in homogeneous detection of proteins and small molecules

Xing Zhou, [a,b,†] *Zhangbo Cui,* [a,†] *Liquan Liu,* [a] *Ziyong Sun,* [c] *Meng Lin,* [a] *Qingyi Hu,* [a]

Hongbo Wang, [*b] *and Xianjin Xiao** [a]

[a] Centre of Reproductive Medicine/Family Planning Research Institute, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, PR China.

[b] Department of Obstetrics and Gynaecology, Union Hospital, Tongji Medical College,Huazhong University of Science and Technology, Wuhan, China, 430022

[c] Department of Clinical Laboratory, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, 430000, China.

† These authors contributed equally to this work

* To whom correspondence should be addressed. Email: xiaoxianjin@hust.edu.cn; hb_wang1969@sina.com

Strand name	Sequence[a]
Probe-B	L-strand: 5'-ACG TAT CAT AGC ATA AG CAA CAC CCT T(-Dabcyl) CC ACT GTA CTT CAT ACA TG-3' S-strand: 3'-TGCATA GTA TCG TAT TC GTT GTG GGA A-5'FAM
M-strand	5'-CAT GTA TGA AGT ACA GTG GAA GGG TGT TGCT-3'
B-strand	5'-CAT GTA TGA AGT ACA GTG GAA GGG TGT T(-Biotin) GCT-3'
C-strand	5'-AGC AAC ACC CTT CCA CTG-3'
D-strand	5'-CAT GTA TGA AGT ACA GTG GAA GGG TGT T(-Digoxin) GCT-3'
Probe-B-E	L-strand: 5' Dabcyl-ACG TAT CAT AGC ATA AG CAA CAC CCT TCC ACT GTA CTT CAT ACA TG-3' S-strand: 3' FAM-TGCAT AGT ATC GTA TTC GTT GTG GGA A-5'
B-strand-E	5'-CAT GTA TGA AGT ACA GTG GAA GGG TGT TGC T-3' Biotin
D-strand-E	5'-CAT GTA TGA AGT ACA GTG GAA GGG TGT TGC T-3' Digoxin
Template	5'-CAGTATATCATAGCAGTATCATAGCATAAGCAACACCCTTCCACTGTACTTCATACATG-3'

Table S1. Sequences used in this work:

[a] Small molecules or fluorescent groups modified on thymines were emphasized in red.

Supplementary modeling and experimental results

1. The influence of labeled biotin alone and the buffer solution of streptavidin on the branch migration process.

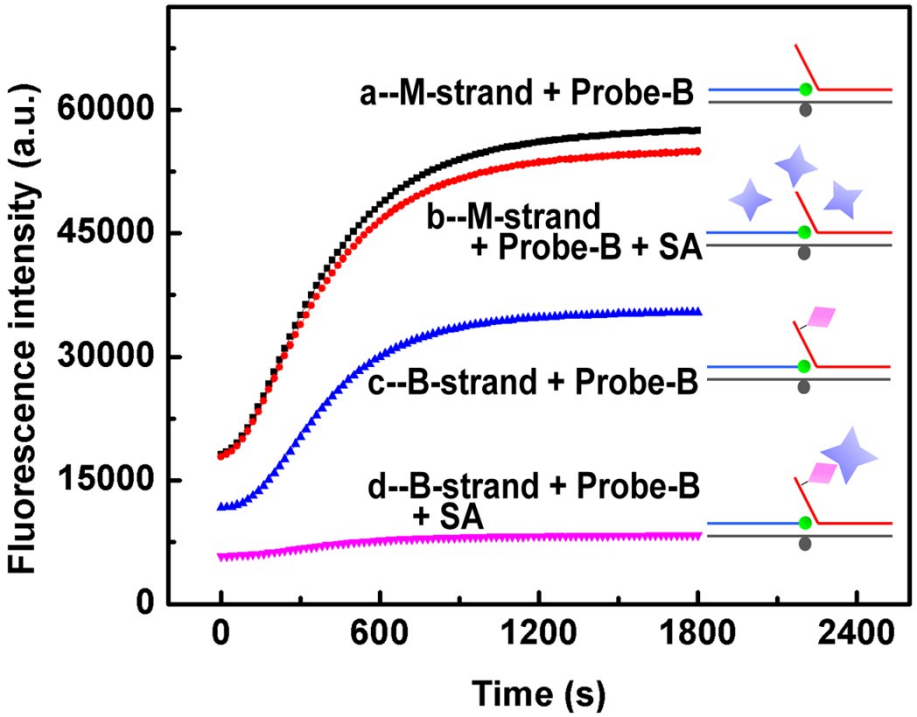

Figure S1. The influence of mid-labeled biotin and the solution buffer of streptavidin on the branch migration process.

2. Thermodynamic model for the small molecule-protein interactions in the branch migration process

The branch migration based strand displacement was depicted in details in Figure 1a, and can be summarized with the following equations:

$$X + S \underset{k_r}{\overset{k_f}{\rightleftharpoons}} I \underset{k_b}{\overset{k_b}{\rightleftharpoons}} J \underset{k_f}{\overset{k_r}{\rightleftharpoons}} L + Y$$

Where, $k_f = 3.5 * 10^6 M^{-1} s^{-1}$ (usually); $k_b = \frac{400}{x^2} s^{-1}$, where x is the length of the branch migration domain); $k_r = k_f * \frac{2}{x} * e^{\Delta G^{\emptyset}/RT} s^{-1}$, where $\Delta G^{\emptyset}$ is the transition free energy of the binding of toe[1-3].

In the design, the length of the toe was larger than 13-nt, so the corresponding $\Delta G^{\emptyset}$ is larger than 80 kJ/mol. So,

$$k_r = k_f * \frac{2}{x} * e^{\Delta G^{\emptyset}/RT} s^{-1} \approx 0$$

Also, $k_f \gg k_b$. Therefore, the kinetics of the reaction of $X + S \rightleftharpoons J$ was quasi-first-order, and after the whole reaction reaching to a quasi-equilibrium state, the vast majority of the species in the solutions were I and J. In our assay, the signal was designed to be proportional to the concentration of J. And the quasi-equilibrium concentration of J and the corresponding conversion ratio of S to J could be calculated from the following equations:

$$\frac{[J]}{[I]} = \frac{k_{bi}}{k_{bj}} = K_b = e^{-\frac{\Delta G}{RT}};\ [S] + [I] + [J] + [L] = c_0;\ [L] = [Y]$$

Where, $[S] \approx 0; [L] \approx 0; \Delta G$ was the transition free energy of the forward branch migration. c_0 was the initial concentration of S. Then, we could calculate the conversion ratio (denoted as C) of S to J as follows:

$$C_{branch\ migration} = \frac{[J]}{c_0} = \frac{e^{-\frac{\Delta G}{RT}}}{e^{-\frac{\Delta G}{RT}} + 1}$$

Took $R = 8.314\ J \cdot mol^{-1} \cdot K^{-1}$, $T = 273.15 + 37 = 310.15K$, we could draw the curve of conversion ratio over ΔG (Figure 1b, curve a). Finally, the inhibition factor could be calculated by the following equation:

$$IF_{branch\ migration} = \frac{C_{1-B}}{C_{2-B}} = \frac{e^{-\frac{\Delta G_1}{RT}}}{e^{-\frac{\Delta G_2}{RT}}} \times \frac{e^{-\frac{\Delta G_2}{RT}} + 1}{e^{-\frac{\Delta G_1}{RT}} + 1}$$

3. Thermodynamic model for the small molecule-protein interactions in the toehold exchange process

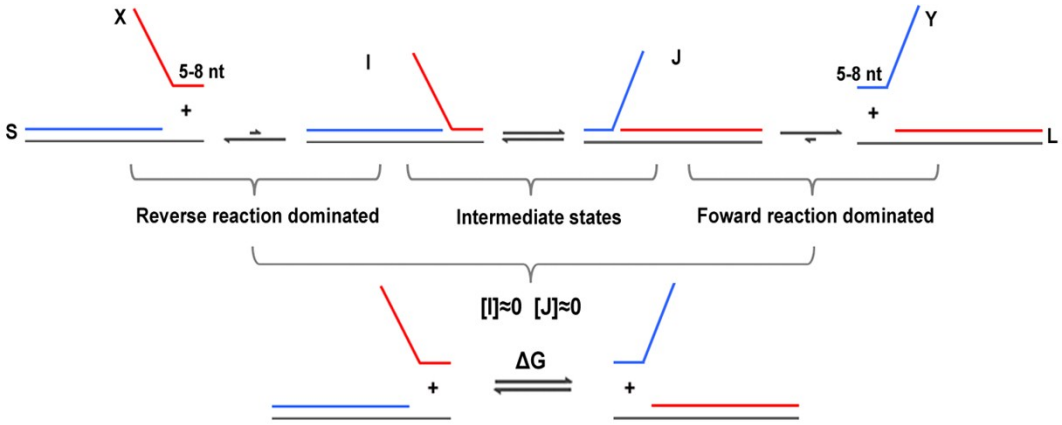

Figure S2. Illustration of the reaction mechanism of toehold exchange based strand displacement.

The detailed branch migration process was depicted in Figure S1, and can be summarized with the following equations:

$$X + S \overset{\Delta G_t}{\rightleftharpoons} I \overset{\Delta G_b}{\rightleftharpoons} J \overset{\Delta G_d}{\rightleftharpoons} L + Y$$

At the equilibrium state,

$$\frac{[L][Y]}{[X][S]} = K_{toehold\ exhcange} = e^{-\frac{\Delta G'}{RT}};\ [S] + [I] + [J] + [L] = c_0;\ [L] = [Y]$$

Where, $\Delta G'$ was the transition free energy of the whole toehold exchange process. c_0 was the initial concentration of S. In toehold exchange process, the length of the toe domain and the dissociation domain was 5-8 nt, and the melting temperature of duplexes of 5-8 bp long was far below 37°C. Therefore, at the equilibrium state,

$[I] \approx 0; [J] \approx 0$

In our design, the signal was proportional to the concentration of Y. Assuming the initial concentration of X and S were both c_0, we could calculate the conversion ratio of X to Y and the inhibition factor with the following equations:

$$Conversion\ ratio = \frac{[Y]}{c_0} = \frac{e^{-\frac{\Delta G}{2RT}}}{e^{-\frac{\Delta G}{2RT}} + 1}$$

$$Inhibition\ factor = \frac{Conversion\ ratio\ (without\ recognition\ protein)}{Conversion\ ratio\ (with\ recognition\ protein)} = \frac{e^{-\frac{\Delta G_1}{2RT}}}{e^{-\frac{\Delta G_2}{2RT}}} \times \frac{e^{-\frac{\Delta G_2}{2RT}} + 1}{e^{-\frac{\Delta G_1}{2RT}} + 1}$$

Where, ΔG_1 was the transition free energy of the toehold process between X and small molecule labeled-S; ΔG_2 was the transition free energy of the branch migration process between X and small-molecule labeled S with bound recognition protein.

4. Calculation of $^{\Delta G_1}$and $^{\Delta G_2}$ for biotin/streptavidin interaction.

According to the main article, $^{\Delta G_1}$ was the transition free energy of the branch migration process between B-strand and Probe-B; $^{\Delta G_2}$ was the transition free energy of the branch migration process between Probe-B and B-strand with streptavidin bound to it. Defining $^{\Delta G_0}$ as the transition free energy of the branch migration process between M-strand and Probe-B, we could obtain the following equations:

$\Delta G_0 = 0; \quad \Delta\Delta G_1 = \Delta G_1 - \Delta G_0 = \Delta G_1; \quad \Delta\Delta G_2 = \Delta G_2 - \Delta G_0 = \Delta G_2$

Where, $^{\Delta\Delta G_1}$ and $^{\Delta\Delta G_2}$ reflected the destabilization effect toward DNA duplexes by small molecule and small molecule/recognition protein complex, respectively.

We then synthesized C-strand, which was 18-nt long and complementary to M-strand and B-strand. Using C-strand, we could extrapolate $^{\Delta\Delta G_1}$ and $^{\Delta\Delta G_2}$ from the following reactions[1-3]:

$$M - strand + C - strand \rightleftharpoons MC\ duplex; \qquad \Delta G_{MC} = -RTlnK_{MC}$$

$$B - strand + C - strand \rightleftharpoons BC\ duplex; \qquad \Delta G_{BC} = -RTlnK_{BC}$$

$$B - strand - SA + C - strand \rightleftharpoons BC - SA\ duplex; \quad \Delta G_{BC-SA} = -RTlnK_{BC-SA}$$

$$\Delta\Delta G_1 = \Delta G_{BC} - \Delta G_{MC}; \qquad \Delta\Delta G_2 = \Delta G_{BC-SA} - \Delta G_{MC}$$

Note: All the above transition free energy corresponded to reactions at 37 °C

To obtain $^{\Delta G_{MC}}$, $^{\Delta G_{BC}}$ and $^{\Delta G_{BC-SA}}$, we measured the melting curves of the above reactions. The melting curves and the melting temperatures were shown in Figure 1c (500 nM), Figure S3 (1000 nM) and Figure S4. For $^{\Delta G_{MC}}$, the calculation process were as follows:

$$K_{MC(337.75K)} = \frac{250nM}{250nM \times 250nM} = 4 \times 10^6 mol^{-1}$$

$$K_{MC(339.55K)} = \frac{500nM}{500nM \times 500nM} = 2 \times 10^6 mol^{-1}$$

$$\Delta G_{MC(337.75K)} = -RTlnK_{MC(337.75K)} = -8.314 \times 337.75 \times \ln\left(4 \times 10^6\right) = -42687 J/mol$$

$$\Delta G_{MC(339.55K)} = -RTlnK_{MC(339.55K)} = -8.314 \times 339.55 \times \ln\left(2 \times 10^6\right) = -40958 J/mol$$

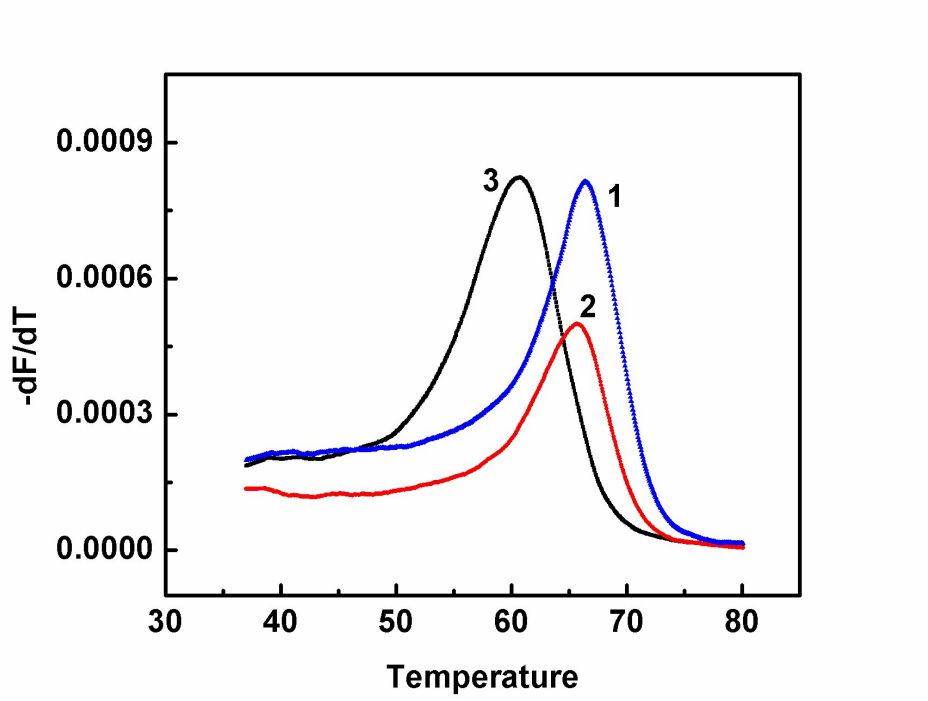

Figure S3. Melt curve peaks of duplexes of M-strand/C-strand (curve 1), B-strand/C-strand (curve 2) and B-strand/C-strand-SA (curve 3).

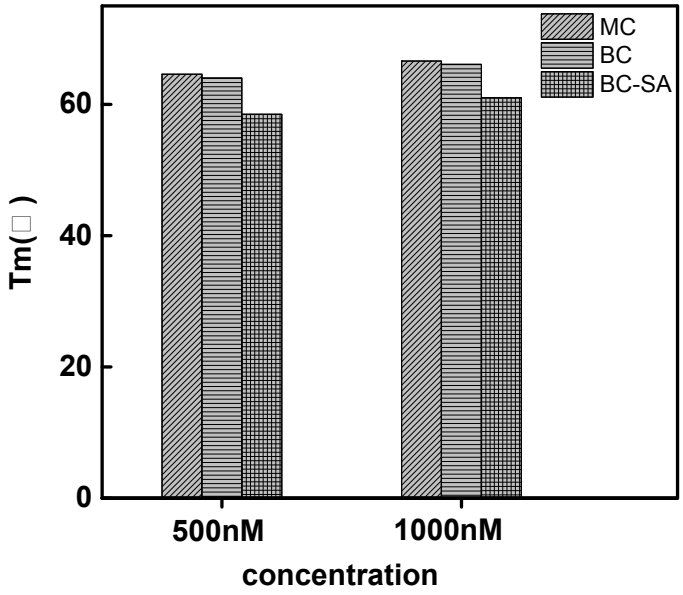

Figure S4. Melting temperatures of BC-SA, BC and MC

Assume that the transition enthalpy change $^{\Delta H_{MC}}$ and the transition entropy change $^{\Delta S_{MC}}$ remained the same values at different temperatures, then

$\Delta G_{MC(337.75K)} = \Delta H_{MC} - 337.75 \times \Delta S_{MC} = -42687 J/mol$

$\Delta G_{MC(339.55K)} = \Delta H_{MC} - 339.75 \times \Delta S_{MC} = -40982 J/mol$

Solving the above equations, and then,

$\Delta G_{MC(310.15K)} = \Delta H_{MC} - 310.15 \times \Delta S_{MC} = -66.218 kJ/mol$

Following the same procedures described above,

$$\Delta G_{BC(310.15K)} = \Delta H_{BC} - 310.15 \times \Delta S_{BC} = -64.335 kJ/mol$$

$\Delta G_{BC-SA(310.15K)} = \Delta H_{BC-SA} - 310.15 \times \Delta S_{BC-SA} = -55.760 kJ/mol$

Finally,

$$\Delta G_1 = \Delta\Delta G_1 = \Delta G_{BC(310.15K)} - \Delta G_{MC(310.15K)} = 1.883 kJ/mol$$

$\Delta G_2 = \Delta\Delta G_2 = \Delta G_{BC-SA(310.15K)} - \Delta G_{MC(310.15K)} = 10.459 kJ/mol$

Using the obtained values of $^{\Delta G_1}$and $^{\Delta G_2}$,

$$IF_{branch\ migration} = \frac{C_{1-B}}{C_{2-B}} = \frac{e^{-\frac{\Delta G_1}{RT}}}{e^{-\frac{\Delta G_2}{RT}}} \times \frac{e^{-\frac{\Delta G_2}{RT}} + 1}{e^{-\frac{\Delta G_1}{RT}} + 1} = 19.1$$

$$IF_{toehold\ exchange} = \frac{C_{1-T}}{C_{2-T}} = \frac{e^{-\frac{\Delta G_1}{2RT}}}{e^{-\frac{\Delta G_2}{2RT}}} \times \frac{e^{-\frac{\Delta G_2}{2RT}} + 1}{e^{-\frac{\Delta G_1}{2RT}} + 1} = 3.4$$

5. Discussion on the deviation between the experimental inhibion factor and the thermodynamic modelling result.

We attributed this deviation to two aspects: First, the themodynamic model assumed that all the signal in our assay was emitted by J, which represented the ending of forward branch migration. However, as the fluorescent group was labeled at the starting site of the branch migration process, the distance between the fluorescent group and the quencher group could gradually increase as the branch migration proceeded. Consequently, partial fluorescent signals were emitted during intermediate branch migration, which produced a fraction of the total signal observed in the presence of streptavidin and lowered down the experimental inhibition factor. Second, the calculation of inhibition factors was based on the value of $^{\Delta G_1}$ and $^{\Delta G_2}$. We

extrapolated the values of ΔG_1 and ΔG_2 through measuring the melting curves of DNA duplexes of MC, BC and BC-SA. The calculation process made several assumptions: i) the DNA duplex in the melting curve analysis involved only two states: hybridized dsDNA state and random coiled ssDNA state. However, this assumption neglected the possible secondary structures of the involved DNA strands[4-6]. ii) In the melting curve analysis, the transition enthalpy ($\triangle$H) and transition entropy ($\triangle$S) were considered to be temperature invariant, which was equivalent to the assumption of temperature invariant heat capacity Cp[7, 8]. Whereas, the Cp of DNA duplex were slightly temperature variant[9, 10], which produced deviation to our calculation. Overall, the discussed two aspects represented the main origins of the systematic deviations and random deviations of our model, respectively.

6. Melting temperatures of end-labeled B-strand/C-strand (E-BC) duplex in the presence and absence of streptavidin (E-BC-SA).

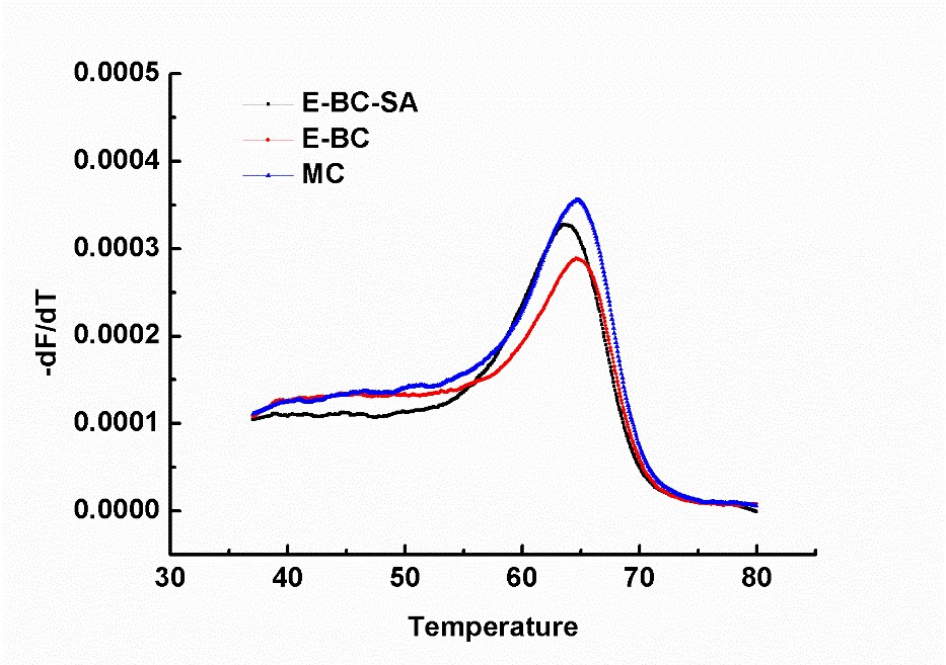

Figure S5. Melting peaks of E-BC-SA, E-BC and MC at concentration of 500nM

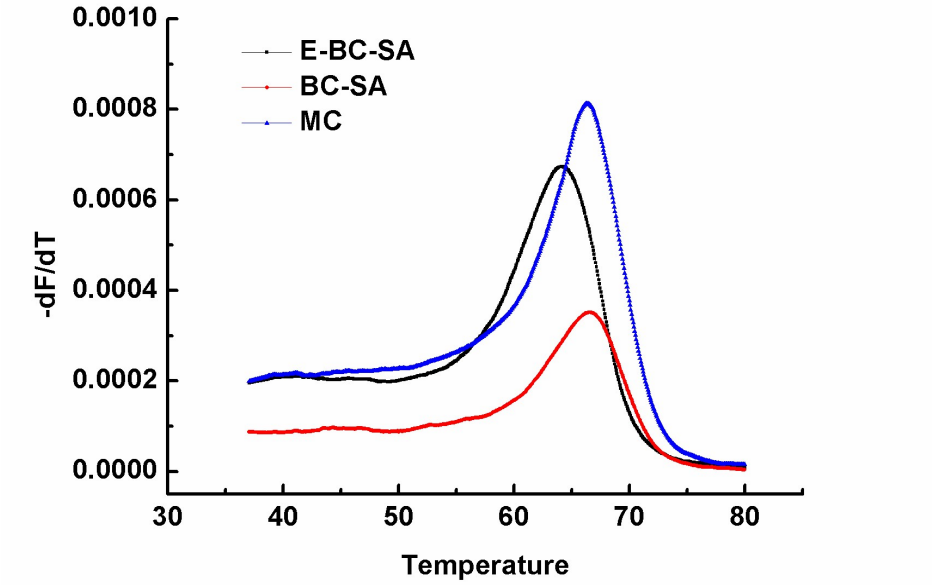

Figure S6. Melting peaks of E-BC-SA, E-BC and MC at concentration of 1000nM

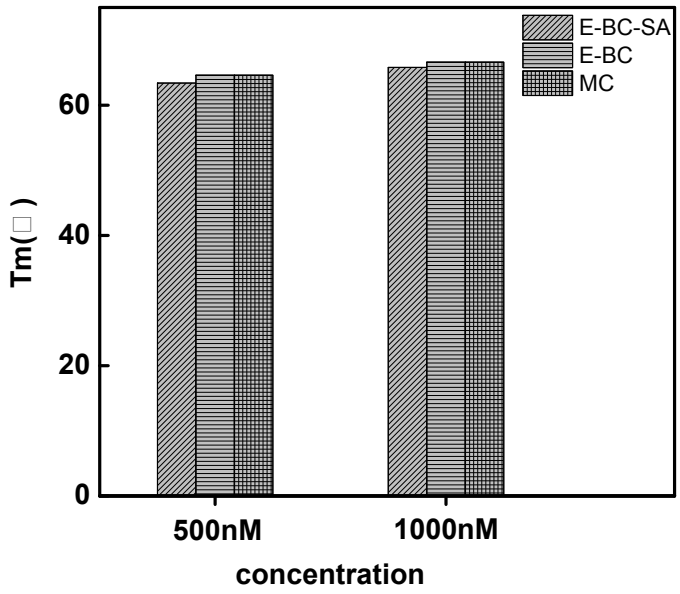

Figure S7. Melting temperatures of E-BC-SA, E-BC and MC

7. Estimation of the binding constant of streptavidin and biotin

The equation for calculation was as follows:

$$K_{net} = \frac{[SB_4]}{[S][B]^4} = \frac{0.25 \times ([B]_0 - [B]_0 \times \frac{F - F_{min}}{F_{max} - F_{min}})}{([S]_0 - 0.25[B]_0 + 0.25[B]_0 \times \frac{F - F_{min}}{F_{max} - F_{min}}) \times ([B]_0 \times \frac{F - F_{min}}{F_{max} - F_{min}})^4}$$

Where, $[B]_0$ was the initial concentration of B-strand; $[S]_0$ was the initial concentration of streptavidin; F was the fluorescence intensity plateau corresponding to a certain concentration of $[S]_0$; $F_{max}$was the fluorescence intensity plateau when no streptavidin was added (Figure S8, curve a). $F_{min}$was the fluorescence intensity plateau when B-strands were completely bound to streptavidin (Figure S8, curve f).

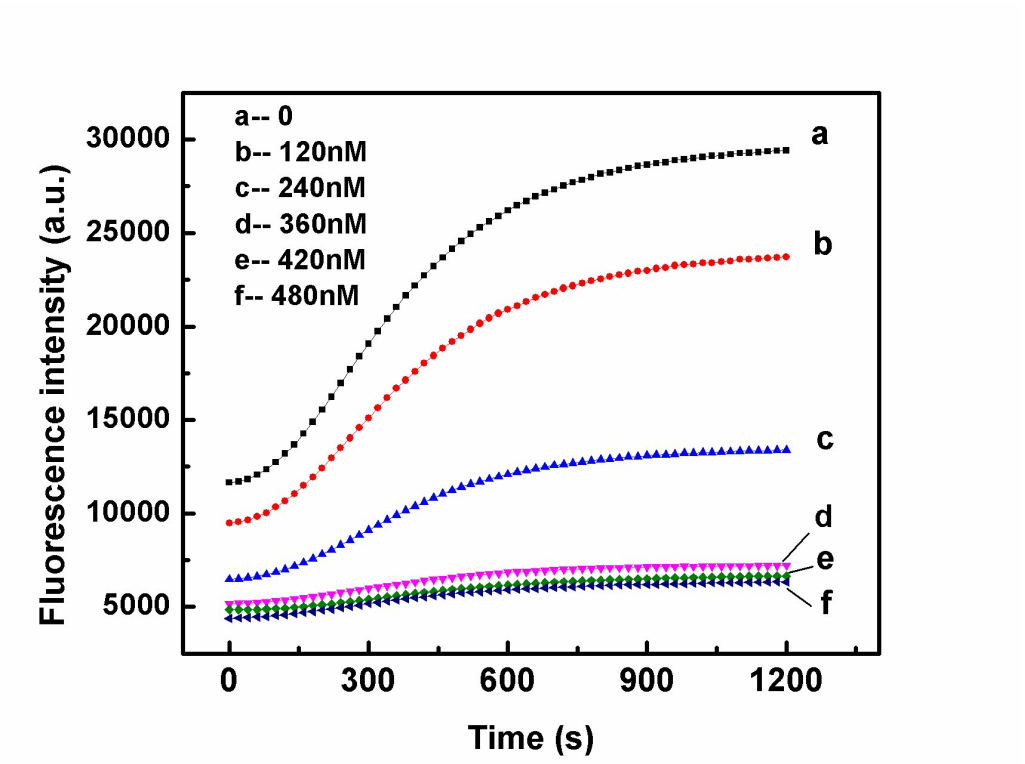

Figure S8. Fluorescence intensity responses of the BM probe system at different concentrations of streptavidin.

We then calculated the values of net binding constant, and results were shown in Table S2.

Table S2. Melting temperatures of E-BC-SA, E-BC and MC

$[S]_0$ / nM	Fluorescence intensity plateau	Calculated K_{net} / M^{-4}
0	29418 (F_{max})	NA
120	23337	3.88×10^{31}
240	13395	2.10×10^{27}
360	7213	3.88×10^{25}
600	6630 (F_{min})	NA

We observed significant variations of the calculated K_{net}. This variation was associated with the complex binding stoichiometry between biotin and streptavidin. As reported in previous literatures[11], the streptavidin was a tetramer, and each monomer had a binding constant of ~10^{14} M^{-1}. Assuming that the four binding events within a single molecule of streptavidin were independent and would not interfere with each other[11], the net binding constant would be ~10^{56} M^{-4}, which was much higher than our calculated values. The vast difference revealed the four binding events would compete for space, leading to a much lower net binding constant. At the equilibrium state, there would be a distribution of SB, SB_2, SB_3 and SB_4. Under different initial concentrations, the distribution would be significantly different, thus rendering distinct net binding constants. We would also like to note that our present assay was not suitable for accurate determination of the binding constant of protein and small molecule, as very slight deviations in the fluorescence measurement would be amplified through mathematics to a large deviation in the calculated binding constant. Nevertheless, our assay was a very convenient tool for estimation and comparison of small molecule-protein interactions.

8. Polymerase elongation after branch migration process

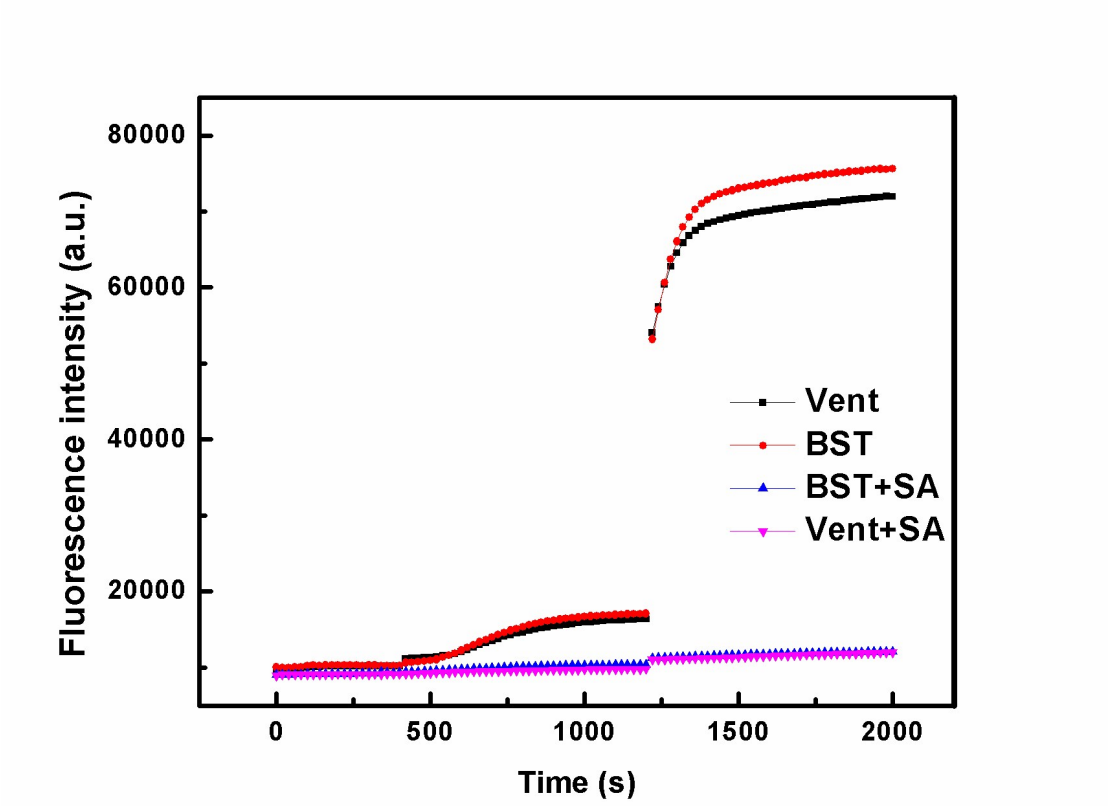

Figure S9. Vent and Bst polymerase elongation after branch migration process. Each solution comprised of 250nM of Probe-B, 200 nM of B-strand, 1 unit of DNA polymerase and 420 nM of SA (if any).

9. Steric hindrance effect of streptavidin on the Bst polymerase elongation process.

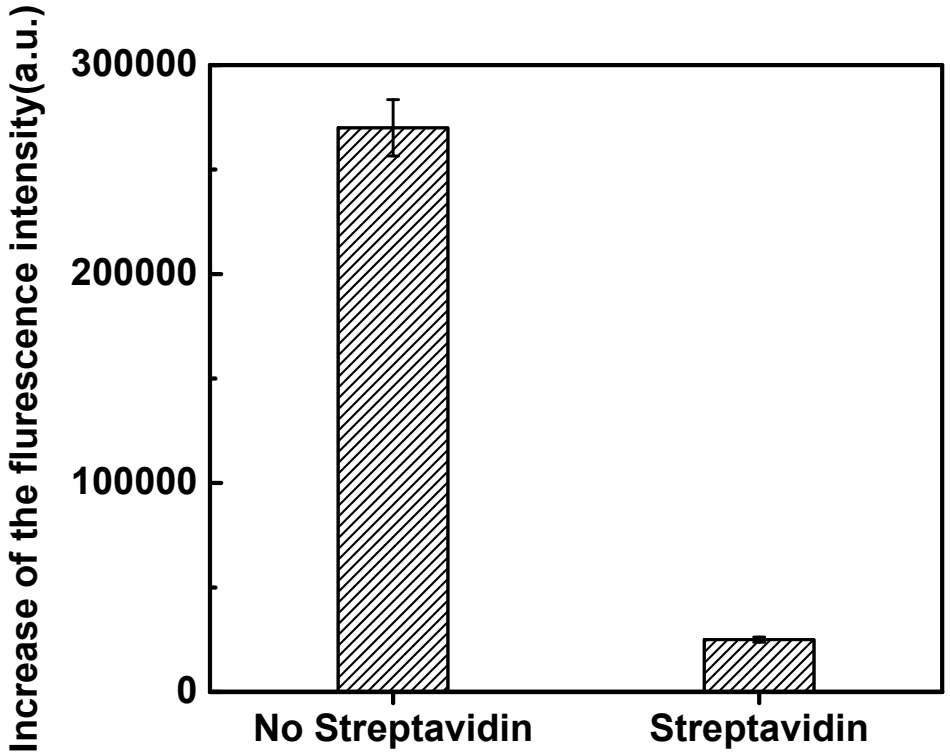

Figure S10. Steric hindrance effect of streptavidin on the Bst polymerase elongation process. The inhibition factor was calculated to be 10. Each solution comprised of 400 nM of B-strand, 400 nM of template, 0.1 unit of Bst polymerase, and 420 nM of Streptavidin (if any). In this experiment, B-strand served as primers.

10. Optimization of the amount of streptavidin

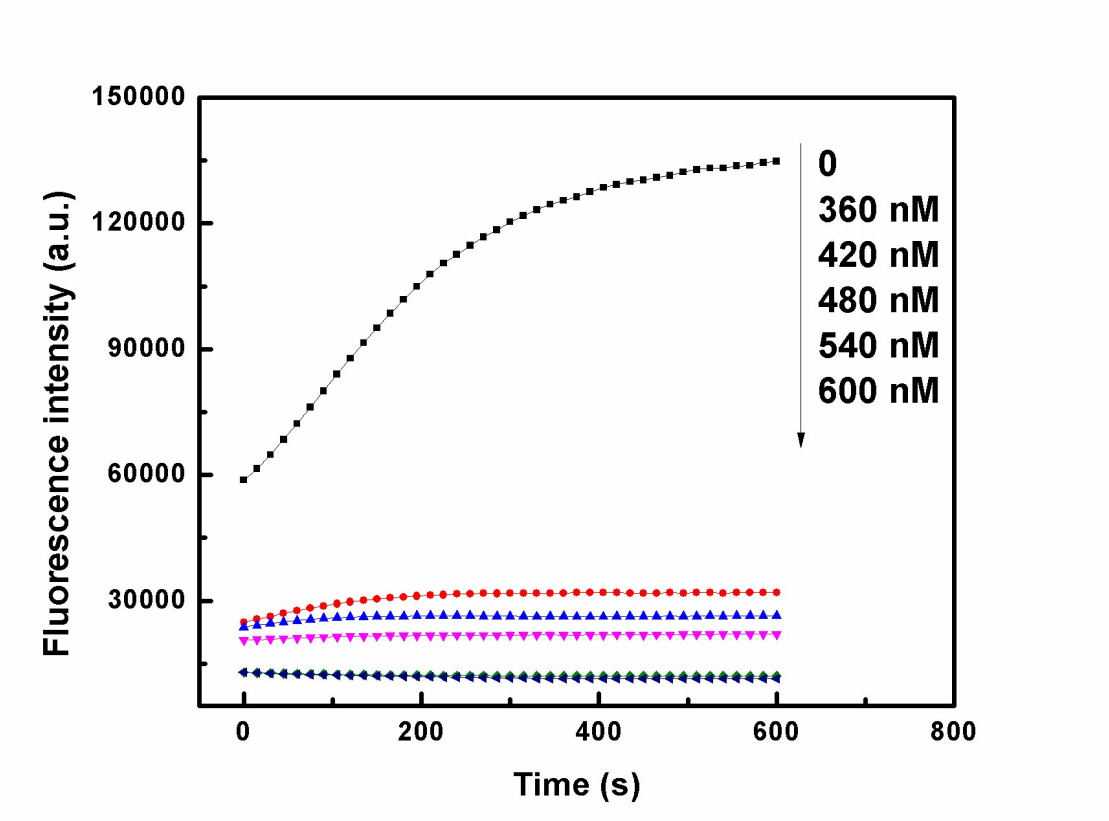

Figure S11. Optimization of the amount of streptavidin based on the increase of fluorescence intensity at stage 3.

The fluorescence intensity decreased as the amount of streptavidin increased. In theory, too little streptavidin would cause high background signal, whereas too much streptavidin could weaken the competition between the labeled biotin and the analyte biotin, and thereby lower down the sensitivity. Based on the above discussion, we chose 420nM of streptavidin as the optimal amount for following experiments.

11. The influence of labelling position of digoxin on the inhibition factors

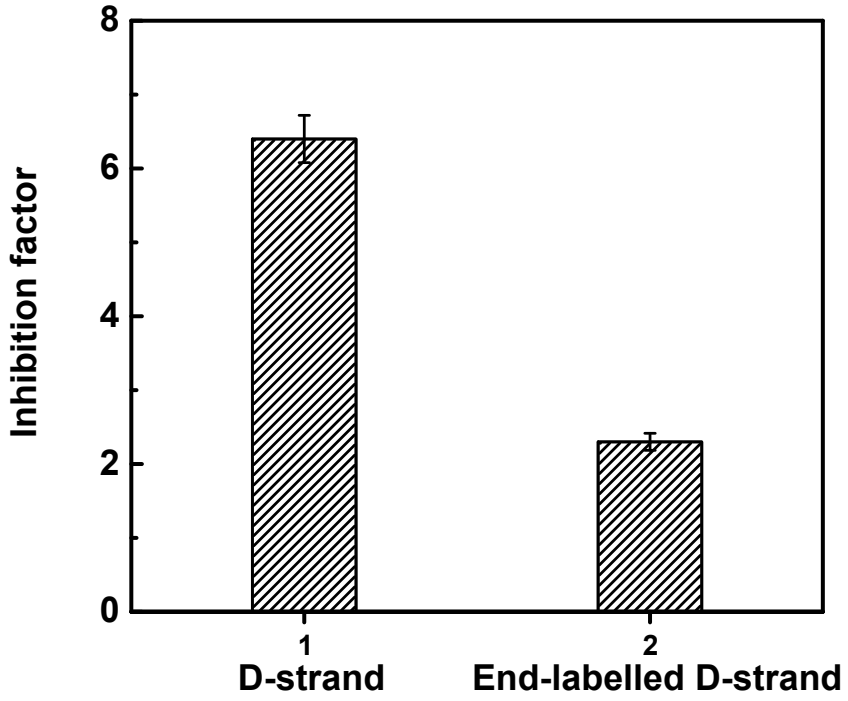

Figure S12. The influence of the labeling position of digoxin on the inhibition factor.

Reference

1. K. J. Breslauer, R. Frank, H. Blocker and L. A. Marky, *Proceedings of the National Academy of Sciences of the United States of America*, 1986, **83**, 3746-3750.
2. J. SantaLucia and D. H. Turner, *Biopolymers*, 1997, **44**, 309-319.
3. R. K. J. Olsen, S. F. Dobrowolski, M. Kjeldsen, D. Hougaard, H. Simonsen, N. Gregersen and B. S. Andresen, *J Inherit Metab Dis*, 2010, **33**, 247-260.
4. M. L. Bochman, K. Paeschke and V. A. Zakian, *Nat Rev Genet*, 2012, **13**, 770-780.
5. G. Bonnet, O. Krichevsky and A. Libchaber, *P Natl Acad Sci USA*, 1998, **95**, 8602-8606.
6. C. Y. Wang, J. H. Bae and D. Y. Zhang, *Nat Commun*, 2016, **7**.
7. I. Rouzina and V. A. Bloomfield, *Biophys J*, 1999, **77**, 3242-3251.
8. I. Rouzina and V. A. Bloomfield, *Biophys J*, 1999, **77**, 3252-3255.
9. P. J. Mikulecky and A. L. Feig, *Biochemistry-Us*, 2006, **45**, 604-616.
10. I. Rouzina and V. A. Bloomfield, *Biophys J*, 2000, **78**, 403a-403a.
11. J. DeChancie and K. N. Houk, *Journal of the American Chemical Society*, 2007, **129**, 5419-5429.