

Response to Reviewer #1 Comments

Thanks for your time reviewing this and giving comments. We have revised the paper according to your valuable suggestions. More caution has been provided regarding the RPKM-relative free energy transformation. As for the conformational entropy, we think that it is indeed the intrinsic cause of the efficiency differences as well as the free energy difference. This measurement is adopted mainly because of the consideration of the simulation-experiment comparison.

Point 1: The Geissler method is Virtual Move MC not Visual Move MC.

Response 1: The typo has been corrected in the manuscript. Thanks for reminding.

Point 2: While the RPKM results must involve free energy of association it is a bit less clear one can use the log ratios to directly produce a relative free energy. The authors were careful about the phrasing but even more caution should be given to the reader.

Response 2: We wholly agree. The RPKM values were used to represent relative free energy since we could not get exact concentration data from the experiments where *E. coli* library genes are simultaneously targeted. We understand that it is not a direct representation of free energy. This method was alternatively adopted to make an experiment-simulation comparison. Much more description was given in the manuscript to illustrate this and provide further caution to the reader about the assumptions made. The revised language

in the results and discussion section is provided below and marked in grey to highlight updated text.

Line 321-326

It is reasonable to believe that these values could reflect the free energy difference, although they are not parameters which are rigorously related to the free energies. Therefore, RPKMs of the cloned copies for one target under the application of different probes were transformed into “free energy differences” by eq(3). This alternative method was adopted so that a quantitative comparison could be conducted between simulation and experiment.

Line 406-407

The imprecise way of obtaining experimental relative free energy might also contribute to these differences.

Point 3: The results in figure 4 are a bit disappointing whereas the enrichment data of figure 1 is clear.

Response 3: This study is truly a first-generation simulation model and was expected to serve as a foundation for future development. We agree that the simulations were unable to accurately capture the experimental results in some cases. We were pleased to see qualitatively that the simulations explained why an extreme long adapter (788nt) did not provide a favourable capture (larger probe-target interaction free energy) and implied why molecular inversion probes (MIP) had inferior performance when capturing larger genes in the *E. coli* open reading frame library. There are several new directions that we are interested in improving after this first experience. The probe-target interaction free energy, according to the simulation results, seems not to be the only factor

that determines the gene-capture efficiency. For instance, the target-probe hybridization free energy may also be a critical factor that contributes to the capture efficiency, which is also the main issue in the following step of our study. We did not expect that this computational study would be able to correctly reproduce the experimental results in all cases, but we hope this was the first step to provide a rationale for why the probe-target interaction free energy plays an essential role in this reaction and, perhaps with a computational simulation, adapters can be designed with this in mind for future gene-capture probes.

Point 4: While free energies were calculated from oxDNA simulations it seems that conformational entropy could be a strong effect in the overall process with respect to length.

Response 4: We agree that conformational entropy should be considered in the computation of the free energies. Our oxDNA calculations directly generate all the conformations of all the adapters and targets. Furthermore, we think that the entropy is a driving factor that leads to the interaction free energy difference. This can be seen in the calculation of the free energies, which can be represented by the equation below.

$$\frac{\Delta G_{int}}{RT} = -\ln \left(\frac{P_{(Q_{ex}=0, Q_{lig}=0)}}{P_{(Q_{ex}=10, Q_{lig}=10)}} \right)$$

The calculation of the conformational entropy corresponding to interacted states, if we adopt the definition of conformational entropy from Andrew J. Doig and Michael J.E. Sternberg, Protein Science, 1995, can be represented as:

$$\Delta S = -R \cdot P \ln P \quad (P \text{ represents } P_{(Q_{ex}=0, Q_{lig}=0)})$$

These functions are both influenced by $P_{(Q_{ex}=0, Q_{lig}=0)}$, the fractional population of the states in which target and probe interact. The difference is that the free energy emphasizes the relative contributions of probe-target interaction and non-interaction states. We agree that the conformational entropy is the intrinsic factor that drives the efficiency differences as well as the free energy differences. Even the mismatch end-to-end distance mentioned in the manuscript could contribute to the low accessibility to the relative states, which is also related to the entropy. Actually, the free energy could also be considered as a conformational free energy that is determined by the entropy difference between non-interaction and interaction states. Nevertheless, the free energy seems to be a more appropriate parameter to be measured if we wish to extract a comparable parameter from experimental data in order to conduct an experiment-simulation comparison. More emphasis regarding the importance of conformational entropy has been added in the beginning of the results part, which is provided below.

Line 259-263

Moreover, conformational entropy may contribute to the performance of the probes. For instance, a longer probe may be ineffective due to the entropic cost of interacting with target sequences. We used computational simulations to understand this issue by estimating the free energy of probe-target interaction, which includes the effects of entropy.

About the improvement of the conclusion: To improve the consistency between the conclusions and results, edits have been made in abstract, result and discussion section, which is provided below.

Line 33-37

Our results suggest that the adapter length is a factor that contributes to the free energy of target-probe interaction, thereby determining the efficiency of capture. The results indicate that LASSOs with extremely long adapters cannot capture the targets well. They also suggest that targets of different lengths may prefer adapters of different lengths.

Line 344-359

The lowest free energy suggests the most favorable option of probes. As the target length increases, a transition in optimal AL from 54-nt (the MIP AL) to 242-nt and 442-nt could be observed. (compare Figure 4A vs Figure 4E, F) The difference between the free energy of MIPs relative to zero free energy reference slightly increases (from 0 to a positive value) when the target length changes from 400 to 1500bp. The increased difference might be a clue as to why LASSOs exhibit more favorable efficiency in capturing and cloning long targets from a computational aspect. The predicted difference, however, is not consistent with the considerable difference between MIPs and LASSOs observed in the experiments. Furthermore, all the simulations showed that the probes with 788-nt adapters demonstrated larger probe-target interaction free energies than the probes with 242- and 442-nt adapters. This result provides a potential explanation for the unsatisfactory performance of 788nt-adapter LASSOs. Along with the experimental results, it suggests that an extremely long adapter will not contribute to improving the efficiency of a gene-capture probe and that the most suitable AL should be within a certain range.

Line 381-388

Although the simulation results captured some of the experimental results and provided a reasonable explanation for some phenomena, such as the failure of 788nt adapters, the computations require further improvement. For instance,

the free energy predictions do not provide a sufficient computational rationale for the inferior performance of MIPs. The considerable interaction free energy gaps between MIPs and LASSOs shown experimentally in the cases of several targets (Figure 4B, E, and F), could not be well captured. The simplicity of the model should be firstly taken into consideration if we start to analyze...