PRITZKER SCHOOL OF MOLECULAR ENGINEERING

THE UNIVERSITY OF CHICAGO

**Andrew L. Ferguson, PhD**

Associate Professor
Vice Dean for Equity, Diversity, and Inclusion

Pritzker School of Molecular Engineering
University of Chicago
5640 South Ellis Avenue
Chicago, IL 60637

p: (773) 702-5950
e: andrewferguson@uchicago.edu
w: www.ferglab.com

September 11, 2021

Prof. Hanadi Sleiman
Associate Editor, *The Journal of the American Chemical Society*

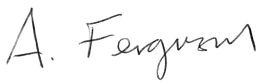
Dear Prof. Sleiman

Thank you for your 7-Sep-2021 message regarding our article "Determining sequence-dependent DNA oligonucleotide hybridization and dehybridization mechanisms using coarse-grained molecular simulation, Markov state models, and infrared spectroscopy" assigned manuscript ID ja-2021-05219p.R1. We thank you for handling our submission and the anonymous reviewers for their careful reading and critical assessment of our work.

We reproduce below the full text of the reviewer comments in blue along with our responses, reproduction of the changes made to our manuscript, and the locations of these edits. The modifications within our revised submission in response to the reviewer comments are indicated in red text.

In light of these changes, we hope that our revised submission is suitable for publication in *The Journal of the American Chemical Society*.

Yours Sincerely



Andrew L. Ferguson, PhD

Associate Professor
Vice Dean for Equity, Diversity, and Inclusion
Pritzker School of Molecular Engineering
University of Chicago

cc: M.S. Jones, B. Ashwood, A. Tokmakoff

## Response to Reviewer # 1

Recommendation: Publish in JACS without change.

Comments: Most of my comments were properly addressed. The manuscript, as far as I understand, is suitable for a journal like JACS.

**Author Reply:** We are pleased that the reviewer is satisfied with our modifications and responses and that they are pleased to support publication in JACS. We would like to thank them again for their time, effort, and expert comments.

## Response to Reviewer # 3

Recommendation: Publish in JACS after minor revisions.

Comments: Jones et al. have updated their manuscript in response to reviewers' comments and improved it significantly. I am essentially happy to proceed towards publication.

**Author Reply:** We are pleased that the reviewer finds the manuscript improved and suitable for publication in JACS subject to minor revision.

Some remaining issues that I would like to bring to the author's attention:

(1) The authors have applied the finite size correction of Ouldridge et al. following a comment by reviewer 2, but it hasn't been applied in the correct way. It would be correct to apply it to the melting curves discussed at the bottom of p7/top of p8. Since the sequences are homodimers, the correction to Tm at the simulation concentration would actually be zero; instead, the widths of transitions would change, as noted by Ouldridge et al.

However, the authors have actually simulated at 3.5x the experimental conc. If the Ouldridge correction - which assumes ideal behaviour - applies, then it should be relatively simple to correct the curves for the concentration difference. Simply divide the phi ratio by 3.5, and then calculate f_hybridised using the modified phi. If the Tms were previously too low, this change will make them lower still (perhaps by a couple of K), but it is the fairest comparison. It shouldn't affect the rest of the analysis, although it is worth noting that the simulations were always performed "at the melting temperature for a conc of 7mM" - there is no well-defined, concentration-independent melting temperature.

However, it is not appropriate to apply the Ouldridge correction to data measuring the dissociation time. The authors effectively measure the statistics of dissociation events in infinite dilution - these are unaffected by the factors that lead to the Ouldridge correction. More specifically, the Ouldridge correction accounts for concentration fluctuations neglected in small, canonical simulations. These concentration fluctuations contribute to the hybridisation rates, not the dissociation rates.

**Author Reply:** We appreciate these valuable comments and have made several corrections and modifications to the revised manuscript in light of these recommendations.

(I) We thank the reviewer for pointing out that the Ouldridge finite-size fluctuation corrections do not affect the dissociation rates and are therefore not relevant for this comparison. We have eliminated our discussion of this correction at the top of p. 16 in Section 3.1 ("To correct for finite-size effects...") and modified the plot and caption in Figure 1 to report the original uncorrected data. As mentioned in our previous response letter, the fluctuation correction led to only a slight change in the reported rates and so this modification does not change any of our analysis or interpretations of the data.

It is true that the homodimer finite-size fluctuation correction does not affect the melting temperature but only widens the transition, and this was indeed what we saw when we applied the correction. We have also edited the text on p. 17 that erroneously suggested that this correction would change the dissociation rate or melting temperature:

"The computational time scales and melting temperatures reported in the remainder of the paper are not corrected by these calibration corrections ~~or finite-size fluctuations~~ since we are only interested in the relative trends in the behaviors of the four sequences."

(II) We appreciate reviewer's suggestion to be more explicit that melting temperatures are a function of concentration, and that our definition applies to a particular model at a particular concentration. As such we have revised the text in several places to note we are using the melting temperature specific to the 7 mM concentration employed in our simulations:

On page 7:

"each sequence was simulated at its respective melting temperature at 7mM concentration as dictated by the 3SPN.2 model"

On page 9:

" MSMs were constructed for each of the four DNA sequences at their respective 3SPN.2 melting temperatures at 7 mM concentration from the 40×25 $\mu$s simulation trajectories"

On page 20:

"Since each system is simulated at the same 7 mM concentration and at its respective melting temperature"

(III) We thank the reviewer for the suggestion to correct our simulation prediction of the melting temperature to the experimental concentration. The proposed strategy is to assume ideal behavior, reduce the computed $\phi$ ratio per Eqn. 7 in Ouldridge et al.[1] by the 3.5× concentration difference, and then recompute $f_{hybridized}$. We believe that this procedure would work under the assumption that the equilibrium constant does not change with temperature, but in general this may not be a good assumption since $-RT_mK_{eq} = \Delta G^\circ = \Delta H^\circ - T_m\Delta S$. Owczarzy et al.[2] engage this issue to develop a van't Hoff-like concentration correction to $T_m$ based on the enthalpies of duplex formation and helix nucleation. In the absence of knowledge of these standard enthalpies, we could instead calculate the melting temperature by performing multiple simulations at a range of temperatures, applying the proposed concentration correction to predict $f_{hybridized}$ at that temperature, then fitting a melting curve from which we could extract the concentration corrected $T_m$. This additional suite of required calculations would be exclusively in service of nailing down the concentration correction and would not change our approach since we intentionally selected the 7 mM 3SPN.2 melting temperature in order to maximize the number of spontaneous transitions between dissociated and hybridized states to provide a rich data set of these events over which to train our MSM.

The approach we take to the concentration correction in the manuscript is an empirical one, wherein we draw a comparison of the computational predictions against experimental measurements. The observed calibration corrections therefore account for net effect of both the 3.5× higher concentration in simulation and the assumptions and approximations inherent in the 3SPN.2 model. As the reviewer observes, the (+4) K calibration correction encompassing both of these effects is overly optimistic since the concentration corrected $T_m$ will be depressed and controlling for this would require a larger positive correction for the error due to the 3SPN.2 model approximations.

In light of the reviewer's suggestion and the discussion above, we have modified the text on p. 16 to alert the reader to these important issues:

"Although the 3SPN.2 model reproduces melting temperatures relatively well, we observed a systematic 4 K under-prediction relative to experiment and so we apply a universal (+4) K corrective temperature shift to our computational results. We note that our simulations are expected to slightly over-predict the melting temperature since they are conducted at 3.5× higher concentration relative to experiment, and the concentration-adjusted corrective shift accounting for the approximations inherent in the 3SPN.2 model would be slightly larger. Owczarzy et al. present an analytical prescription to apply concentration corrections to the melting temperature using knowledge of the enthalpy changes associated with duplex formation and helix nucleation[2]. In the absence of these quantities, one could instead empirically estimate concentration corrected melting temperatures by conducting a suite of simulations over a range

of temperatures and assume ideal molecular behavior to apply concentration corrections to the observed hybridized fractions[1]. (As demonstrated by Ouldridge et al., finite-size corrections do not affect the melting temperature for homodimers[1].) Furthermore, we note that these empirical calibration factors to the 3SPN.2 predictions are applied only for the purposes of making an experimental comparison, but acknowledge that there are uncertainties introduced by assuming that the equilibrium dynamics at fixed temperature can be compared directly to relaxation kinetics following a 15°C T-jump."

(2) I realise now that the criterion for dissociation is that both of the central pairs separate to 1.3nm, which is more reasonable than my previous understanding. However, although the authors say that the estimates of kdŝlow are robust to a range of choices for this cutoff, the yellow, red and green curves all have a noticeable shift between 1.3 and 2.0nm (unsurprisingly blue does not, since the GC pairs are in the core). So what exactly does robust mean? Why not just report the data for larger cutoffs?

**Author Reply:** Our original motivation for selecting a 1.3 nm cutoff was to maintain consistency between that for fraying and dissociation. This is, however, an unnecessary constraint since these are fundamentally different processes. As the reviewer recommends, we now report the data in Figure 1 using the larger 2.0 nm cutoff. We do indeed see relatively small changes in our results upon changing the cutoff from 1.3 nm to 2.0 nm, producing an average change in the calculated values of $k_d^{\text{slow}}$ of only 7%, demonstrating that our calculation is robust to the particular choice of cutoff within this range. This modification does not affect any of our subsequent analysis or interpretation.

We have updated the text on p. 15 as follows:

"First, we tracked the slow response corresponding to duplex dissociation in our simulations by compiling the distribution of times at which both of the central base pairs first separate to a distance of 2.0 nm starting from an initial fully hybridized duplex. This cutoff was selected as the distance beyond which the strands are effectively non-interacting and defines the dissociated state. We extracted our computational estimate of $k_d^{\text{slow}}$ by fitting a decaying exponential to the fraction of hybridized sequences as a function of time $f_{\text{hybridized}}(t) = \exp(-k_d^{\text{slow}} t)$. We verified that our cutoff was sufficiently large by observing that our calculated values for $k_d^{\text{slow}}$ changed by an average of only 7% by adopting a 1.3 nm cutoff."

(3) In their response to point (7) in my original review, the authors talk about "averaging" over all microstates. Do they mean estimating F as

F = -kT ln sum_i exp(-F_i/kT)

where i runs over all microstates in the macrostate? This is very different from averaging F_i, as might be inferred from the text.

**Author Reply:** Thank you for pointing out our lack of clarity on this point. We have modified the text on p. S10 of the Supporting Information to more clearly convey the details of this calculation:

In applying the NN model to each macrostate, we made the simplifying assumption that the ensemble of microstates constituting each macrostate could be represented by a single pattern of Watson-Crick base pairing that are schematically illustrated in Fig. 2a. In reality, each macrostate is composed of an ensemble of microstates that may have slightly different base pairing patterns and explicitly averaging over all of these microstates may change the predictions of the NN model. To evaluate the extent to which the microstate ensembles would change our free energy estimates, we calculated a NN microstate free energy by taking a simple average over the NN free energies predicted for each configuration within a given microstate – a reasonable assumption given the structural similarity of configurations within a microstate – then computing macrostate free energies via a weighted average according to the MSM equilibrium occupancy probabilities for each microstate. These operations can be represented as follows, where $i$ indexes over the $N_j$ configurations in microstate $j$, $j$ indexes over microstates, and $p_j$ represents the MSM equilibrium occupancy of microstate $j$,

$$F^{micro} = \frac{1}{N_j} \sum_{i \in micro} F_i^{NN} \tag{1}$$

$$F^{macro} = \sum_{j \in macro} p_j F_j^{micro} \tag{2}$$

Recalculating the NN predictions in this way leads to changes in the $\Delta F = F - F_H$ values reported in Fig. 3 (i.e., the free energy of each macrostate relative to the hybridized state) of <1 kJ/mol. As such, representing each macrostate by a single representative microstate and base pairing pattern is a highly accurate simplifying approximation for the purposes of this calculation.

**Response to Reviewer # 4**

Recommendation: In between publish elsewhere and accept without change.

Comments: The authors have addressed the comments of all the reviewers very thoughtfully and thoroughly. Although I'm still of the opinion that it's probably just below the level for JACS, I do think it is a very good paper, and wouldn't argue against acceptance if that's the majority view of the reviewers.

**Author Reply:** We thank the reviewer again for their time and effort in reviewing our work, and are pleased that they found all comments to have been thoroughly addressed.

PS I'm glad the 7M was just a typo!

**Author Reply:** Us too!

## References


[1] Ouldridge, T. E.; Louis, A. A.; Doye, J. P. *J. Condens. Matter Phys.* **2010**, *22*.

[2] Owczarzy, R.; Vallone, P. M.; Gallo, F. J.; Paner, T. M.; Lane, M. J.; Benight, A. S. *Biopolymers: Original Research on Biomolecules* **1997**, *44*, 217–239.