Revealing Thermodynamics of DNA Origami Folding via Affine Transformations

Jacob M. Majikes, Paul N. Patrone, Daniel Schiffels, Michael Zwolak, Anthony J. Kearsley, Samuel P. Forry, J. Alexander Liddle.

1 Physical Measurement Lab, National Institute of Standards and Technology, Gaithersburg, MD, 20899-6203, USA

*J. Alexander Liddle Tel: 301-975-6050; Email: liddle@nist.gov

Present Address: J. Alexander Liddle, 100 Bureau Drive, Gaithersburg, MD 20899-6203, USA

ABSTRACT

Structural DNA nanotechnology, as exemplified by DNA origami, has enabled the design and construction of molecularly-precise objects for a myriad of applications. However, limitations in imaging, and other characterization approaches, make a quantitative understanding of the folding process challenging. Such an understanding is necessary to determine the origins of structural defects, which constrain the practical use of these nanostructures. Here, we combine careful fluorescent reporter design with a novel affine transformation technique that, together, permit the rigorous measurement of folding thermodynamics. This method removes sources of systematic uncertainty and resolves problems with typical background-correction schemes. This in turn allows us to examine entropic corrections associated with folding and potential secondary and tertiary structure of the scaffold. Our approach also highlights the importance of heat-capacity changes during DNA melting. In addition to yielding insight into DNA origami folding, it is well-suited to probing fundamental processes in related self-assembling systems.

INTRODUCTION

The sequence specificity inherent to DNA self-assembly provides nanofabrication options orthogonal to those of existing top-down methods, enabling the construction of multifunctional, addressable 3D nanostructures. Nanoparticles, quantum dots, fluorophores, biomolecules, and combinations thereof can be positioned with single-nanometer precision and accuracy on DNA platforms. Recent work has demonstrated production at scale, facilitating the manufacture of drug delivery vehicles, calibration artefacts, and large area, functional surfaces at low cost. However, while raw material scale-up is necessary for the industrial implementation of DNA nanofabrication technology, it is not sufficient. As observed in Moore’s seminal 1965 paper, the semiconductor industry has seen success in part because "No barrier exists comparable to the thermodynamic equilibrium considerations that often limit yields in chemical reactions." If artificial macromolecular self-assembly is to replicate the success of the semiconductor manufacturing paradigm, at a minimum we must fully understand those limiting thermodynamic equilibria and be able to predict the nature of the resulting defects and their populations. This will enable the development of design, process, and assembly strategies that ensure functionality.
Unfortunately, DNA nanostructure thermodynamics and kinetics are difficult to characterize via the techniques typically employed by biophysicists, as are similar complex cooperative effects relevant in biological DNA. DNA origami (15) – in which a long, single-stranded “scaffold” folds by binding many “staple” oligomers – has been the focus of much work on yield and folding. Most efforts, to date, have been theoretical, (16, 17) indirect, (18, 19) or unable to resolve individual folding events. (20–23) In addition, with few exceptions (24–26), these studies have examined structural yield, i.e., fidelity with respect to a designed shape, as opposed to functional yield, i.e., inclusion of all necessary active components. The large number of potential design configurations means that measurement techniques must be adapted to nanofabrication contexts, and they must also have sufficient throughput to adequately explore an extensive parameter space.

In this work, we investigate the effect of fold distance and persistence length on the thermodynamics of an initial folding event using high-throughput, quantitative Polymerase Chain Reaction (qPCR) equipment. These measurements are enabled by our modular fluorescent melt-curve reporter system, which minimizes extraneous signals, in combination with an affine transformation technique that significantly improves melt-curve baseline and background correction by iteratively collapsing numerous replicate measurements of a sample onto a single curve through linear parameters representing known sources of variation (27). While this study is motivated by a desire to understand the details of the origami assembly process, the methods and results are applicable to numerous types of DNA-based self-assembling systems.

Characterization of folding events in DNA origami is problematic due to the sheer number of hybridization events. One example origami structure, the Rothemund tall rectangle (15), consists of a 7.2 kilobase M13 bacteriophage viral scaffold bound together with ≈ 200 synthetic staples, each of which binds two or three different and disjoint sub-sequences of the scaffold (15). It must be emphasized that this results in ≈ 600 reversible hybridization reactions, of typically 6 to 24 bases in length, with varying degrees of independence, collectively responsible for ≈ 200 forced links in the topology of the scaffold. Design choices include scaffold sequence permutation, staple motifs, and scaffold routing – all of which influence the folding process. Attaining the imaging resolution required to observe individual binding events is a significant challenge and has, so far, only been performed post mortem, on assembled structures (25). To develop an understanding of design-assembly relationships, functional staple yields must be characterized for numerous “equivalent” structures of varying design under various processing conditions.

These design-assembly relationships for DNA origami are a function of the interplay between several cooperative energetic contributions (17, 25). The primary contributor to hybridization comes from base stacking and base pairing as described by the nearest-neighbor model (28). Additional contributions come from coaxial end-to-end base stacking with adjacent staples, and entropic effects, as sub-sequence binding events change the topology of the scaffold (16, 17, 28). As each staple binds, it reduces the entropic penalties for further folds and provides base stacking for its neighbors. It also changes the equilibrium constant for the other sub-sequences on that staple from a bimolecular to a unimolecular form. While this is not the focus of this paper, we note that the complex problem of
cooperative folding can be addressed using our programable test system. Here we only explore zeroth order cooperative effects, i.e., changes in conformational entropy. While the thermodynamics and kinetics of biological DNA hybridization (28–30) and the basic theory for loop entropy in short (tens-of-bases) ssDNA (28) and in long (kilobase) dsDNA (31, 32) are all established, they are not directly applicable to DNA origami. The latter begins as long ssDNA loops, likely with intrinsic secondary structure, which are folded into smaller and smaller mixed ssDNA/dsDNA loops until the entire structure comprises dsDNA helices bound together by Holiday junctions. Neither the loop entropy model for short ssDNA nor that for long dsDNA is sufficient to predict conformational entropy in these complex, dynamic systems of mixed persistence length. Additionally, the large thermal processing range for DNA nanofabrication (approximately 25 °C to 80 °C), as compared to normal physiological temperatures (36.5 °C to 37.5 °C), has an important consequence: the temperature dependence of the enthalpy and entropy, expressed as the change in heat capacity on melting, \( \Delta C_p \), cannot be neglected. While the significance of \( \Delta C_p \) for DNA melting is established (33, 34), a general consensus on typical values and sequence dependence is lacking, and it is commonly neglected entirely in secondary structure prediction (28, 35). It is worth noting that \( \Delta C_p \) shifts the hybridization energetics in proportion to the difference between the evaluation temperature and the melt temperature and therefore becomes critical over the broad temperature ranges used in nanofabrication.

So far, we have discussed the importance of and need for a better understanding of the thermodynamics of DNA origami assembly. Below, we examine the limitations of current techniques when applied to this problem. Experimental characterization of DNA origami systems with existing biological techniques is often as fraught as application of existing theory. For smaller ssDNA loops, it is common to use UltraViolet-Visible (UV-Vis) spectroscopy to measure melt curves or differential scanning calorimetry (DSC) to directly measure the melting energetics (36). In both cases, a significant mass of DNA is required, and both techniques measure signal from all DNA in the sample. Sufficient resolution to distinguish one approximately 80-base staple or, worse, an approximately 24-base staple sub-sequence, from the other 7.2 kilobases of DNA in the system is not yet available. For measurement of cyclization of long dsDNA, ligation reactions are often quenched with Ethylenediaminetetraacetic acid (EDTA), then gel electrophoresis is used to quantify the relative concentration of cyclized/uncyclized material for sequential snapshots of the reaction (37). In the case of origami, there is little to no gel mobility shift associated with a single staple fold of the scaffold. Intercalating dye fluorescence may be used to determine melt temperatures, (38) but there is a lack of consensus regarding baseline correction techniques (39) to which thermodynamic parameter extraction, via van't Hoff analysis (33, 40, 41) plotting \( \ln(K) \) vs \( 1/T \), is exceptionally sensitive. As a result, the melt temperature, \( T_m \), is taken to be the only reliable parameter, and is often the only one reported (42). The variant of van't Hoff analysis, in which \( T_m \) is plotted against DNA concentration, common to both fluorescence and UV-Vis, is not usable for scaffold folding itself. This is because, whether a fold occurs due to the formation of secondary structure between disparate domains of the scaffold, or as the result of a second domain of an already-bound staple hybridizing, the change in topology is a unimolecular process, requiring
neither the addition (removal) of another strand to (from) the structure. Therefore, the concentration terms cancel out in the equilibrium constant.

qPCR equipment, which enables high-throughput fluorimetry, has been used with intercalating dyes to measure anneal/melt curves of whole origami(16, 17, 19), while Förster resonance energy transfer (FRET) pairs have been used to measure local melting of origami(43), inter-origami base stacking(44), and DNA tile assembly(45). However, absent a robust baseline correction method, these techniques cannot provide quantitative information. Baseline correction is difficult because fluorophores, whether FRET pairs or intercalating dyes, have a temperature-dependent fluorescence efficiency that is also sensitive to their local environment. This includes neighboring ssDNA/dsDNA transitions, as well as the DNA sequence immediately adjacent to the fluorophore. However, if the issues associated with local environment effects can be mitigated, and the baseline correction made rigorous, the specificity and sensitivity of FRET reporters makes them attractive for measuring folding events. Here, we overcome these difficulties by implementing a modular reporter system that maintains a uniform fluorophore environment for every fold measured and introduce an analysis approach that alleviates baseline correction issues. In this analysis we make, and validate, the assumption that we can separate the effect of the variables subject to error from the intrinsic signal. In other words, the physics responsible for FRET, i.e., the relative proximity of donor to acceptor, is the same for all samples, while the measured signal depends on the sample volume, well location, detection/excitation efficiency, alignment of equipment optics per well, etc. This assumption enables us to apply an affine transformation (see Eq. 2) to the melt curve and significantly reduce subjectivity in the baseline correction process, replacing it with a more mathematically rigorous approach. With a reproducible fluorescence signal, and a robust baseline correction method in hand, we make use of the high throughput of qPCR to explore the large parameter space covering the effects of fold distance and persistence length.

MATERIAL AND METHODS

Test system:

Using the design capabilities of structural DNA nanotechnology, we created a minimal, single-fold test system to observe the effects of loop entropy without the occurrence of numerous simultaneous events which can significantly obfuscate analysis.

An illustration of the measured folds is shown in Fig. 1a. This system is coincidentally similar to those being used for subattomole detection of microRNA(46). Despite important distinctions between the thermodynamics of these systems, both make a strong case for the utility of simplified DNA self-assembly systems. We examined the melting of an identical sequence of DNA in a simple bimolecular no-fold system as well as with 13 different folding distances dividing the loop into figure-eights. The system which does not fold the scaffold will be referred to as the hybridization-only system. An important feature of this system is that we may vary the effective, or apparent persistence length, $L_p$(47–49). This is critical since, for the same length of biopolymer, $L_p$ controls the effective number of freely-jointed units in the polymer chain, which in turn controls the number of possible configurational
microstates and the entropic folding penalty. The apparent persistence length also determines the minimal fold distance below which enthalpic bending costs become a significant factor. Since the scaffold is circular, we only probe folding distances up to half the scaffold length.

![Diagram](image)

Figure 1 a) folding transition schematics for bimolecular hybridization-only (H.O.), i.e., staple binds to scaffold, and unimolecular, i.e. bound staple hybridizes fully, fold reactions. b) Persistence length strands: name, anticipated $L_p$, and schematic of oligomers/scaffold. c) Reporter scheme: cartoon of fold and reporter strands on the scaffold and description of sub-sequences for fold and reporter strands where X' indicates the complement to $X$. d) Relative positions of melt and anchor sub-sequence complements on M13 scaffold.

We adjust the apparent persistence length, $L_p$, by varying how much of the scaffold is either dsDNA and ssDNA. To systematically do so the scaffold was divided into 37 base segments; two sets of complimentary oligos for each segment were ordered, namely a 37-base full complement and a 32-base complement omitting the last 5 bases at the 3' end. The scaffold $L_p$ was then varied by repeating motifs of these complement strands as shown in Fig. 1b. The short and long $L_p$ systems contained only the 32 and 37-base complements respectively, while the medium system alternated between the 32 and 37-base complements across the scaffold. In addition, a native system with no complements, and a blocked system with only 32-base complements in areas of high predicted secondary structure, SI section 1, were examined. For each $L_p$ system, a pool of complement strands was created for all segments except those complementing the anchor, melt, and fluorophore sequences for that fold. This pool was added in 5x excess of the scaffold.

While the $L_p$ of ssDNA and regularly-nicked dsDNA(50) have been measured, determining the persistence length of these composite constructs is beyond the scope of this work, particularly since there is some controversy regarding the persistence length of short dsDNA segments (51–54).

However, we provide a rough estimate in SI section 3, for later use in allowing us to implement configurational entropy and bending enthalpy models.
Our reporter system must meet two criteria to effectively measure and compare fold distances. First, it must measure melting of the same sequence of DNA for each fold, to simplify comparison. Second, the local environment around the FRET pairs should be as uniform as possible across samples. To satisfy these constraints, we used the same fluorophore/quencher-labeled oligomers for all folds, both to minimize sequence-dependent changes in fluorescence as well as variability in concentration associated with measuring concentration and pipetting multiple strands.

One potential drawback of this approach is that the reporter strand does not create the Holiday junction-like structure found in typical origami. However, since the structure formed is the same between the hybridization-only control and the test samples, it has no effect on the excess energies due to changes in fold distance or persistence length.

Our reporter scheme is illustrated in Fig. 1c. The reporter comprises three strands and the scaffold, with 4 domains/sub-sequences within those strands. The melt, quench, and fluorophore sub-sequences were identical sequences for all samples, while the presence/complementary location of the anchor sub-sequence in the fold strand changed between samples to enforce fold distance. The lengths of these sub-sequences were chosen so that the melt sub-sequence $T_m$ is more than 10 °C below that of any other sub-sequence. The right side of Fig. 1d illustrates the set of folding distances. For all 12 replicates of each folding distance at each persistence length, the sample was pipetted, mixed, annealed, and finally melted while measuring fluorescence intensity. Further details, such as the sample preparation scheme to minimize unwanted errors and fully capture pipetting variation are addressed in the methods.

**Data analysis: baseline and background correction**

The signal in optical melt curves, whether from FRET, UV-VIS or intercalating dye measurements, does not solely measure the completion of a hybridization reaction. The source of the signal, whether absorbance or fluorescence, is temperature dependent both in the pre- and post-melt states. Additionally, various other contributors to the measured signal, such as sample/plate autofluorescence\(^\text{55}\) and pipetting error, will lead to variability in background intensity between replicates of the measured curves. Inherently, the methods used to remove both temperature dependent baseline and background signal will change the data in the melt transition, making rigorous correction critical.

Baseline correction of optical melt curves typically consist of a two-window function fit across user-selected windows for the high and low temperature shoulders of the sigmoid\(^{40, 56}\). These fits are then extrapolated across the transition of interest, and the measured signal is corrected via Eq. 1, where $h(T)$, $l(T)$, $m(T)$, and $\sigma(T)$ are the high-temperature plateau fit, low-temperature plateau fit, measured signal, and intrinsic signal, respectively.

$$\sigma(T) = \frac{m(T) - l(T)}{h(T) - l(T)}$$

[1]

The appeal of the standard two-window fit is its simplicity and nominal ability to address both temperature dependence and background signal. However, this approach is only valid if the fit function is truly representative, if the fit window does not include any points from the transition, and if the sources of measured signal variation only shift or magnify the melt curve rather than distorting it.
While these assumptions may be true for UV-Vis measurements, they are problematic for fluorescence measurements. Although examples of non-linear baseline functions have been used (39), it is not clear that any of these functions, linear or otherwise, accurately represent the complex mix of factors contributing to the baseline fluorescence. Additionally, the final analysis result can be highly dependent on user-selected fitting windows, especially if there is noise in the measurement. Given that there are not clear objective figures of merit for this window selection, the subjective nature of the two-window fitting approach creates significant uncertainty.

To address these shortcomings, we apply an approach based on affine transformations. Critically, we do not assume the functional form of the baseline or background. Rather, we assume that the physical processes contributing to the baseline are the same for every replicate of a sample, but we allow the scale of the intrinsic signal to be a variable, as a result of, e.g., pipetting error. In other words, we assume that the measured signal is related to the intrinsic signal through linear combinations of a finite number of physical processes, whose relative contributions are determined as part of the analysis. With a sufficient number of sample replicates, the figure of merit for optimization in the analysis is how well all replicate melt curves collapse onto a single intrinsic melt curve via affine transformation, as shown in Fig. 2(27).

Unprocessed melt curves, acquired as described in the methods, are shown in Figure 2a. The samples share the same characteristics, i.e. melt temperature and functional form, arising from the intrinsic signal. Affine transformations were applied to collapse these curves and reveal the intrinsic signal in Figure 2b.

A detailed description of this process is included in SI. Section 2. In brief, we first use a set of empty wells to determine the plate autofluorescence and optics alignment-related background, $B(T)$. Next, we use a set of wells containing all strands, bar the folding strand, to determine the fluorescence-temperature-dependent baseline, $R(T)$, are subtracted from a) to b), but are separately measured in control experiments.

Figure 2. Hybridization-only (H.O.) control for the short Lp scaffold before a), and after b), affine transformation. The empty plate background, $B(T)$, and fluorescence temperature dependence baseline, $R(T)$, are subtracted from a) to b), but are separately measured in control experiments.
single intrinsic melt signal, $\sigma(T)$ [Eq. 2]. In all such cases, the relative error on the collapse is a useful metric for the uncertainty of the intrinsic melt signal.

As the transformations are applied directly to experimental data we never need to, nor do we, specify functional forms for the baseline or background. This frees us from making poor modeling choices about the actual behavior of $\sigma$, $R$, and $B$, aside from needing to satisfy an equation having the general form of Eq. 2. For this analysis, an additional optimization constraint is imposed, specifically requiring a monotonic slope at the high temperature plateau. While there are enough data points to represent the entire melt curve, there were insufficient low-temperature points on all curves to impose an additional low-temperature monotonicity constraint.

$$m_i(T) = a_i\sigma(T)R(T) + b_iB(T) + c_i$$  \[2\]

For the analysis presented here, we assume that FRET quenching reduces the fluorescence intensity of the folded state to a sufficiently low level that its temperature dependence is negligible. This may not be true for every system and buffer condition.

One subjective choice in the affine transformation is application of a monotonicity constraint, and the window over which it is applied. The data presented here, Fig 7 in particular, was analyzed with only a high temperature monotonicity constraint. There are insufficient data points on the low temperature plateau of the control and long fold samples to apply a low temperature monotonicity constraint without some sensitivity to window choice. This is illustrated in SI section 12.

Eq. 2 may be easily modified for other systems. For example, for intercalating dyes the free-dye fluorescence temperature dependence cannot be neglected. Additionally, terms for free-dye fluorescence would be supplemented with terms for free-dye fluorescence in the presence of ssDNA. These possibilities are discussed in our best-practices paper for this analysis(27).

One limitation of this method is that the curve-collapse optimization is most effective when only one physical process dominates the measured signal.

As a counter-example, if variations in pipetted volume contribute a similar change in signal as the background, the optimization can become inefficient or even infeasible. Additionally, while this approach removes much subjectivity from the analysis, the choice of initial equation, e.g. Eq. 1, and the linearity constraints at high and low temperatures are still choices for the experimentalist.

Ultimately, implementation of affine transformations in analysis of optical melt curves allows us to simultaneously perform baseline and background corrections with reduced subjective, model-choice errors.

Data analysis: thermodynamics extraction

Calculation of $\Delta H$ and $\Delta S$ is performed via van’t Hoff analysis; in which the equilibrium constant, $K$, is calculated and fitted to Eq. 3 as a function of temperature.

$$\ln(K) = \frac{\Delta H}{RT} + \frac{\Delta S}{R}$$  \[3\]

Equilibrium constants are calculated using Eq. 4 & 5 for the hybridization-only control (bimolecular) and folded samples (unimolecular) reactions, respectively. Where $[ssDNA]$ and $[dsDNA]$ refers to the concentration of the scaffold in the melted, or dehybridized, and the annealed, or hybridized, state respectively, while $[excess]$ refers to the relative excess of the fold strand to the scaffold. Each of
these is treated as a fraction multiplied by the total scaffold concentration, which in Eq. 5 cancels out, as one would expect a unimolecular reaction to be concentration independent. As previously described, this is because the reporter signals when the melt sub-sequence of the fold strand hybridizes, Fig. 1c, at which point the more stable anchor sub-sequence of that strand is already bound to the scaffold. As such, the folding (unfolding) of the scaffold neither consumes (generates) a free staple strand, and is effectively unimolecular in nature. The exception to this is the hybridization-only control, which does not fold the scaffold and is bimolecular

\[
K = \frac{[ssDNA][foldStrand]}{[dsDNA]} = \frac{[ssDNA][excess–dsDNA]}{[dsDNA]} = \frac{[ssDNA][excess–1+ssDNA]}{[1–ssDNA]} \tag{4}
\]

\[
K = \frac{[ssDNA]}{[dsDNA]} = \frac{[ssDNA]}{[1–ssDNA]} \tag{5}
\]

Estimating ΔS requires extrapolation to \(1/T=0\), which is to say, infinite temperature. Small uncertainties in the extrapolation of \(\ln(K)\) as a function of \(1/T\), give rise to large uncertainties in the value of the intercept at \(1/T = 0\), i.e., ΔS. While a linear relationship between \(\ln(K)\) and \(1/T\) is often assumed in the van’t Hoff analysis, the change in heat capacity, ΔC\(_p\), results in a nonlinearity. The relationship between ΔC\(_p\), ΔH(T) and ΔS(T) is described in Eq. 6.

\[
\Delta C_p = \Delta H/\Delta T = T \cdot \Delta S/\Delta T \tag{6}
\]

\[
\Delta T = T_{\text{ref}} - T_m \tag{7}
\]

\(T_{\text{ref}}\) an arbitrarily chosen temperature at which \(\Delta H\) and \(\Delta S\) are evaluated. It is also worth noting that fitting ΔC\(_p\) from curvature in van’t Hoff analysis of optical melt curves is not always successful(35). Limited data sampling, in units of \(1/T\), makes the typical fitting of the curvature in that space highly dependent on the quality of the data. For a thorough discussion of this, we direct the reader to the relevant section in the review by Mikulecky and Feig(33).

**Loop entropy and worm-like chain bending model**

The change in entropy resulting from converting a linear polymer to a loop can be modelled using the Jacobson-Stockmayer entropy extrapolation(28). This may be readily rearranged, SI section 3, to give an estimate of the loop-to-figure-eight transition entropy shown in Eq. 8. Where, in number of bases of ssDNA, \(N_0\) is the length of the original loop, and \(N_1\) and \(N_2\) are the lengths of the two-resulting figure-eight portions illustrated in Fig. 1. \(\Delta S_{\text{ref}}\) is the entropy lost in forming an ssDNA loop comprising C nucleotides. This value was experimentally determined for small hairpins, and is extrapolated to larger loops. In our predictions we use the values for a 30 nucleotide hairpin(28). The relationship between ΔS and the j-factor terminology used for long dsDNA cyclization is shown in Eq. 9.

It is worth noting that the units for C are effectively “nucleotides of ssDNA.” The value of C, together with the prefactor, empirically represent the effective number of bases per segment length. Applying Eq. 8 to DNA nanofabrication systems requires a unit conversion from bases of the scaffold systems
shown in Fig. 1 to bases of ssDNA. The tentative conversion factors we used are discussed in SI section 3.

\[ \Delta S_{\text{fold}} = \Delta S_{\text{ref}} + 2.44R \ln \left( \frac{N_1 N_2}{30N_0} \right) \]  

\[ \Delta S_{\text{fold}} = -R \ln(f) \]  

Implementation of the worm-like chain (WLC) model for the energy of bending, \( E_b \), per unit length of the arc length of the polymer is given in Eq. 10, where \( R \) is the radius of curvature of the bend(57). As such the \( \Delta H \) associated with bending is given by Eq. 11, where \( L_0 \) is the contour length of the DNA.

\[ E_b = \frac{L_p k_b T}{2R^2} \]  

\[ \Delta H_{\text{bend}} = \frac{L_p k_b T}{2R^2} \ast L_0 = \frac{L_p k_b T}{2(\lambda_0/2\pi)} \ast L_0 \]  

As noted above, our mixed dsDNA/ssDNA system is complex, and a detailed microscopic treatment, beyond the scope of this paper, would be needed to accurately describe its mechanical properties. We therefore emphasize that Eqs. 8-11 serve only as a rough guide to the relevant energetics.

**FRET Reporter and folding strands**

As noted above, it is vital to maintain a uniform environment for the fluorophores for all measured systems, including DNA sequence. We achieve this by implementing the fluorophore-quencher system illustrated in Fig. 1c, in which the sequences adjacent to the fluorophore (JOE), and the quencher (Iowa BlackF), do not change. The folding distance (Fig. 1a) is controlled by a 42-base anchor sequence that is adjusted to bind at different locations around the loop of the scaffold strand. To minimize non-fold-related fluorescence background, we control the stoichiometry of our system (Table 1) to ensure that, at low temperature, all fluorophores are bound to a scaffold resulting in a consistent local environment.

To explore the effect of zeroth-order topology changes on melting behavior, we use 13 different folding distances, controlled by the binding location of the anchor sequence (Fig. 1d). The fluorophore, quencher, and anchor sub-sequences all have a predicted \( T_m \) of \( \approx 70 \, ^\circ\text{C} \) under our buffer conditions as evaluated by typical nearest neighbor models(30, 58). These temperatures are well above the predicted 35 \(^\circ\text{C}\) \( T_m \) of the melt sub-sequence, ensuring that the fluorophore is unquenched only when the melt sub-sequence dehybridizes.

Buffer conditions were chosen to match standard DNA nanofabrication protocols: 50 mMol/L Cacodylate buffer, 12.5 mMol/L Mg Acetate, 1 mMol/L EDTA, pH 7.2. Cacodylate was chosen as it buffers pH well across a wide temperature range. DNA sequences are given in section 11 of the SI.

**Persistence Length**

The systems with different apparent persistence lengths illustrated in Fig. 1b are the native scaffold, short (repeating 32 base dsDNA interspersed by 5 bases ssDNA or free scaffold), medium (repeating 32 bases dsDNA, 5 bases ssDNA, 37 bases dsDNA, followed by 32 bases dsDNA), and long (repeating 37 bases dsDNA separated by nicks) This varies \( L_p \) between that of ssDNA (\( L_p \approx 1 \, \text{nm} \) to 3
nm(59)) and regularly nicked dsDNA ($L_p \approx 14$ nm to 43 nm depending on counterion charge and concentration(50)).

The lowest persistence length, ssDNA native M13 scaffold, will in practice contain a great deal of secondary structure. We identified some scaffold locations which could form > 10-base secondary structure across long scaffold distances (SI. section 1), with $T_m$ on the same order as that of the melt sub-sequence. As such, we included an additional scaffold system to investigate the effect of this secondary structure. In this system, we introduce a number of “blocking” strands, designed to hybridize with and out compete this long-distance secondary structure.

We denote the variants of our scaffold, in order of persistence length, as “native”, “blocked”, short, medium, and long. A description of the blocked system and potential secondary structure is given in SI section1, while the sequences for strands creating different apparent persistence lengths, $L_p$, are given in SI section 14.

**Melt experiments**

Our experimental matrix consists of 13 fold distances across scaffolds with 5 different persistence lengths. In addition, it is desirable to replicate each configuration a sufficient number of times to assess the variation due to pipetting while also identifying and removing spurious data sets such as those compromised by bubbles in the wells. We therefore choose to perform 12 replicates of each experiment to ensure high-quality data.

The melt protocol consisted of an initial denaturing step (80 °C for 1 min) an annealing sequence (75 °C to 15 °C, 0.61 °C steps, 3.5 min hold) and finally melting (15 °C to 75 °C, 0.61 °C steps, 3.5 min hold) in which the fluorescence was measured. The raw fluorescence intensity in the green channel, rather than the calculated multicomponent intensity, was used, as uncertainties in the instrument’s multicomponent intensity calculation were unknown.

<table>
<thead>
<tr>
<th>Absolute</th>
<th>Relative excess</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Concentration</strong></td>
<td><strong>(to scaffold)</strong></td>
</tr>
<tr>
<td>Scaffold</td>
<td>45 nMol/L</td>
</tr>
<tr>
<td>Fold Strand</td>
<td>225 nMol/L</td>
</tr>
<tr>
<td>$L_p$ Strands</td>
<td>225 nMol/L</td>
</tr>
<tr>
<td>Quencher</td>
<td>450 nMol/L</td>
</tr>
<tr>
<td>Fluorophore</td>
<td>30 nMol/L</td>
</tr>
</tbody>
</table>

*Table 1. Strand concentration, absolute and relative to scaffold.*

Samples were distributed over two plates for each persistence length set. These plates consisted of 12 replicates of each fold distance, 12 replicates of a fluorescence baseline control containing no fold strand, and 24 empty background wells.

To ensure a consistent sampling of pipetting error simultaneously with manageable sample preparation, 12 replicates of all buffer and strands except the fold strand were independently prepared. These master replicates were used as a base stock for the 12 replicates of each fold distance by multichannel pipettor through the two plates for each scaffold persistence length. As such,
pipetting variability within any fold distance is uncorrelated, but pipetting variation is correlated between the Nth replicate of all folding distances. The qPCR equipment had a certified temperature precision of 0.31 °C, and an optical shot noise of approximately 50 to 100 intensity counts with experiments measuring an intensity change of ≈ 0 to ≈ 20,000 intensity counts.

Differential Scanning Calorimetry

DSC scans were performed on equipment with a 300 µL active volume, baseline repeatability of 0.028 µWatts, and short-term noise of 0.015 µWatts, courtesy of the University of Maryland/NIST Institute for Bioscience & Biotechnology Research (IBBR). While the buffer conditions were identical to the melt experiments performed in qPCR equipment, the strand concentrations and excesses were modified to obtain sufficient signal. Specifically, strand concentrations were much higher and were kept at a 1:1 ratio to avoid measuring the formation of secondary structure in excess strands.

To identify the annealing peaks, scans were performed on subsets of the reporter with increasing numbers of strands. Initial scans were run with equimolar melt and truncated scaffold strands at 90 µMol/L. This sample was removed from the DSC, measured, and equimolar fluorophore strand was added for a final concentration of ≈ 75 µMol/L. The sample removal was repeated, and the quencher strand was added for a final concentration of ≈ 62 µMol/L. Reuse of the same sample allowed confirmation of individual peak identities.

This final sample containing equimolar concentrations of all strands was then pipetted to a 96 well plate and run in the qPCR for comparison. Unfortunately, running a full 10 µL, the minimal sample volume, at the same concentration as the DSC sample would create signal well outside the dynamic range of the detector in the qPCR unit, which would result in significant measurement error. As dilution could significantly change the melt temperature, we chose to take 0.76 µL of DSC sample and pipetted it into 9.25 µL of mineral oil (a mixture of alkanes and cycloalkanes) in the PCR plate. As the mineral oil is not miscible with the water, and floats to form a capping layer, this allowed us to measure the DSC sample without losing water to evaporation. All scans were performed from 5 °C to 100 °C at a rate of 1 °C/min.

RESULTS
**DSC confirmation of fluorescent reporter:**

We use DSC as an orthogonal measurement technique to confirm that the fluorescence reporter system is measuring the anticipated reaction. We do this using a reduced system that allows direct comparison between DSC and fluorescence measurements.

Given that the scaffold is approximately seven hundred times more massive than the 10-base melt sub-sequence, a DSC sample with sufficient melt sub-sequences to provide measurable signal would strain solubility, in addition to being exorbitant in cost. It would also be difficult to analyze due to the additional noise associated with 7.2 kilobases of DNA. DSC was therefore run against a synthetic model system identical in sequence to the reporter, but with a truncated scaffold 100 bases in length. Additionally, the fold strand contained no anchor sub-sequence. All sequences and fluorophores were otherwise identical. In other words, the model system differs from the bimolecular hybridization-only system by omission of 7.15 kilobases of scaffold. Details of the DSC scans are addressed in the methods.

### Table 2. \( T_m \) for reporter from Nearest-Neighbor, DSC, and Fluorescence. Uncertainties are one standard deviation derived from three repeat runs on each sample.

<table>
<thead>
<tr>
<th></th>
<th>( \alpha )</th>
<th>( \beta )</th>
<th>( \delta )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nearest-Neighbor Model</strong></td>
<td>54.9 °C</td>
<td>79.8 °C</td>
<td>80.9 °C</td>
</tr>
<tr>
<td><strong>DSC</strong></td>
<td>51.0 ± 2.3 °C</td>
<td>81.2 ± 0.5 °C</td>
<td>84.7 ± 0.3 °C</td>
</tr>
<tr>
<td><strong>Fluorescence</strong> (maximum of ( \delta F/\delta T ))</td>
<td>54.4 ± 0.25 °C</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

The results of DSC scans are shown in Fig. 3. From a) through c), the sample was removed from the DSC and an additional strand was added, as indicated by the inset schematics in Fig. 3. As the strands were added sequentially, we can readily identify the peaks associated with the dissociation of the sub-sequences illustrated in Fig. 1c: \( \alpha \) (melt), \( \beta \) (fluorophore), \( \gamma \) (quencher). Table 2 shows the calculated \( T_m \) for each sub-sequence. As strands are added the shift in the \( \alpha \) peak and loss of the secondary shoulder apparent in Fig. 3a versus Figs. 3b and 3c suggests that there is some unintended interaction between the nominally non-binding regions of the fold and scaffold strands, enhancing the stability of the intended fold-scaffold duplex. The shifts in both \( \alpha \) and \( \delta \) peaks between Figs. 3b and 3c are likely due to minor base stacking between the melt and fluorophore strands.

The \( T_m \) shown in Table 2 display a reasonable match with nearest-neighbour model predictions. The progression of, and the peaks in Figs. 3a-c, compared to the fluorescence measurements shown in Fig.
3c (red), confirm that the FRET reporter is probing the intended melt sequence illustrated in Fig. 1c. It is reasonable to infer that the fluorescence melt curves measured here are coming from the melt sub-sequence, α, rather than from some spurious side reaction.

$\Delta H$ extracted from the DSC data shown in Fig. 3 is addressed in SI section 4. These values are not discussed here due to the significant uncertainty introduced by the ambiguity of the DSC baseline correction arising from the close spacing of the peaks.

**Melt experiments**

The melt curves shown in Fig. 4a, demonstrate the low noise resulting from affine transformation across 12 replicates. By design, the melt sub-sequence responsible for these melt curves does not change. As such, all differences between these samples should result from the shift between unimolecular and bimolecular reaction types, entropic folding penalties, and potential enthalpic helix-bending penalties, where the latter two result from, e.g., different conformational microstates. The trends shown in Fig. 4a, for the short persistence length, match with thermodynamic intuition. The hybridization-only sample is less stable than the others as it involves a bimolecular reaction, at low concentration, rather than a unimolecular reaction. For the folded samples, increasing fold distances shift the melt curve to lower temperatures indicating an increasing entropy change on melting, as
expected. The error bars shown in Fig. 4, and all other figures presented, represent one standard
deviation above and below the mean of the replicates.

Figure 4b shows van’t Hoff plots for the folded samples. The hybridization-only sample is not
shown here as its equilibrium constant is calculated for a bimolecular reaction, and is
therefore outside the plotted range. The van’t Hoff plots exhibit a slight curvature associated with
ΔC_p.

The T_m from these curves is reported on a semi-
log_2 plot in Fig. 5. While changes in T_m are specific
to this system, they effectively illustrate the energetics associated with these changes in fold
distance and L_p.

It is apparent from Fig. 5 that the melt temperature
of the native scaffold systems generally occurs at
a higher temperature than the other L_p systems. This is consistent with the varied and extensive
secondary structure expected in the native scaffold. Due to this local secondary structure, in
the form of hairpins, the native scaffold may have
much smaller effective looping lengths than the
other L_p systems, resulting in a lower entropy
penalty on folding.

To examine whether the controlling secondary structure comprises many relatively short dsDNA regions
or a small number of long-distance folds, an additional “blocked” scaffold system was tested. In this,
regions of potential long-distance intra-scaffold folds were blocked with ssDNA oligos, SI section 1.
This system follows the same trend as the native scaffold, suggesting that the long-distance secondary
structure is not responsible for the increased T_m. However, the blocked scaffold has a higher T_m for the
large folds, suggesting that some competition with native secondary structure has been removed, as intended, by the introduction of the blocking staples.

Excepting the shortest fold for the most rigid scaffold, the remaining samples follow two clear trends. Increasing loop distance depresses the \( T_m \), as does increasing persistence length, primarily at short fold distances. The 138-base fold for the fully dsDNA scaffold has only 4 jointed sections of dsDNA interspersed with nicks. As such, it may be that the low number of sections allows for some unusually stable conformation(60) and a \( T_m \) higher than expected, a feature also seen for the short and long persistence lengths.

As it relates to DNA origami, the melt curves in Fig. 4, and the melt temperatures in Fig. 5, tell an interesting story which corroborates those developed in previous work(16–18, 21). The relatively low melt temperature for the hybridization-only control, combined with the temperature increase for unimolecular reactions suggests that full origami folding follows a three-step process, although confirming this process will require additional testing in systems with multiple folds. First, the long staple sub-sequences (≥ 16 bases) would bind through bimolecular reactions. Second, the free ends of these partially-bound staples would hybridize in a unimolecular reaction. Third, these folds should dramatically increase the rate of subsequent unimolecular folding events by effectively decreasing the remaining fold distances. We note that, as can be seen in Fig. 5, the increase in melt temperature due to a short versus a long fold distance can be > 10 °C. As this difference is due to an increase in \( \Delta S \), the change

![Melt Temperature vs Loop Distance](image_url)

**Figure 5.** \( T_m \) as a function of the distance along the scaffold between the fold and melt subsequences (\( \log_2 \)) and system \( L_p \). H.O. indicates the hybridization-only control. The single error bar represents the 0.3 °C temperature uncertainty as supplied by the manufacturer. The scatter in loop distance was added to the plot to allow the individual \( T_m \) values to be visualized.
in $T_m$ will depend on the magnitude of $\Delta H$ and/or $\Delta S$ for that reaction, and will be more pronounced for shorter sub-sequences.

van’t Hoff analysis & Heat Capacity Corrections

Fig. 6 illustrates the van’t Hoff analysis of optical melt curves, and the effect of neglecting $\Delta C_p$ terms. It is worth reiterating that extraction of $\Delta C_p$, and $\Delta H$ and/or $\Delta S$ temperature dependence are equivalent actions as per Eq. 6.

The curvature in the van’t Hoff plot can be subtle, making extraction of reliable values for $\Delta C_p$ problematic. This is exemplified by earlier work (61) which indicated that it was not possible to extract $\Delta C_p$ from optical melt curves. Our ability to do so in this case can be attributed both to our affine transformations as well as to improvements in equipment and experimental protocols since the 1990s.

While there is a debate in the literature regarding whether it is acceptable to neglect $\Delta C_p$ in thermodynamic parameter prediction/extraction(33, 45), we consider these results to be evidence of the non-negligible effect of $\Delta C_p$ in DNA melting.

Additional considerations regarding fitting choices, such as fixing individual thermodynamic parameters during fitting, or the role of pipetting errors on van’t Hoff analysis uncertainties, are discussed in SI sections 5-9. Entropy-enthalpy compensation is addressed in SI section 10, where we give evidence that the apparent compensation in this system is likely due to neglect of $\Delta C_p$(62).

The results of van’t Hoff analysis for the short Lp system are shown in Fig. 7. Here, we use a three-parameter fit, and $\Delta H$ and $\Delta S$ are reported referenced to 44 °C, as this is approximately the mid-point of the $T_m$ range. A discussion of this choice of convention is in SI section 13. A benefit of this choice is that it better represents the uncertainty in the system. The error in the fit is amplified as the reference temperature diverges from the range over which the data is fitted.

The trends reported in Fig. 7 all agree with thermodynamic intuition. For $\Delta H(T)$, above some critical maximum curvature one would expect to pay an enthalpic penalty for bending the helix. One would expect this to bending term to be large initially, and to diminish quickly with increasing loop size. For $\Delta S(T)$, one would expect a long-distance fold to reduce the number of configurational microstates more than a short fold. Overall, we would not anticipate $\Delta C_p$ to change with topology to a measurable extent,
and this is what we observe, within error, and the values are consistent with the wide range of per-base ΔCₚ values found in the literature\(^\text{(33)}\).

The composite construct is much stiffer than ssDNA, and calculating its behaviour is non-trivial. This affects the estimation of both ΔH(T) and ΔS(T).

For excess ΔH(T), the composite will have a higher enthalpy of bending at a given radius of curvature than ssDNA. Unfortunately, a rigorous model for this composite system is significantly beyond the scope of this work, particularly for folds with few composite repeat units, where the ssDNA and dsDNA segments would have to be treated independently.

In order to estimate the excess ΔH(T) we used the bending energy from the Worm-Like-Chain (WLC) model and the L_p of dsDNA, 50 nm\(^\text{(57)}\). This choice requires a caveat for two reasons. First, the short fold distances, where excess ΔH(T) is highest, comprise only a small number of composite repeat units, limiting the applicability of the WLC model. Second, the L_p of these composite constructs at lengths long enough for this continuum model to apply is unknown.

To estimate ΔS(T), we use the Jacobson-Stockmayer extrapolation for ssDNA, but we need to account for the fact that the same length of ssDNA, as compared to our composite, will contain many more Kuhn lengths and thus be able to sample many more configurational states. We use the Kuhn lengths of ssDNA and dsDNA to derive an appropriate conversion factor, SI section 3, assuming that the number of repeat units is sufficiently large to behave statistically.

The dotted lines in Fig. 7 are derived using these approximations, and therefore serve only as an indication of the energetic effects that should be borne in mind when considering the observed trends. Further details of the approximations are discussed in section 3 of the SI.

However, even with these limitations, our extracted ΔH(T) and ΔS(T) show similar trends compared to our estimates. It is thus reasonable to interpret the dip in ΔH(T) relative to the hybridization-only control
at short folds as being consistent with a bending energy penalty. Similarly, it is reasonable to assign the increase in $\Delta S(T)$ as a function of fold distance to configurational entropy costs.

We note here that, while the melt curves for medium and long $L_p$ systems, shown in SI section 8, appear by eye to be of equally high quality as that of the short $L_p$ scaffold, their slightly higher uncertainty propagates aggressively through the van’t Hoff analysis. This, combined with the relatively minimal difference in anticipated $L_p$, prevent us from being able to evaluate trends between the various $L_p$ systems in the same way we could for melt temperatures (Fig. 5).

The complexity of predicting both $\Delta H(T)$ and $\Delta S(T)$ as a function of topology change in a partially folded DNA origami will likely require sufficiently nuanced modelling to capture not only changes in folding, but also changes in $L_p$ associated with single staple-subsequence binding events. The uncertainties in $\Delta H(T)$, $\Delta S(T)$, and $\Delta C_p$ are correlated, and their variations as a function of $L_p$ and fold distance are small compared to the hybridization-only control. Higher resolution data is needed to detect these subtle effects, but would require equipment with a temperature control precision better than 0.3 °C, and with a data sampling density greater than one measurement per 0.6 °C.

The improvements provided by the affine transformation and the high number of replicates is dramatic, allowing us to fit $\Delta C_p$ then correct for it, a task often not possible with optical data(61). Even though the uncertainties in $\Delta S(T)$ are inherently greater than those in $\Delta G(T)$ (62), we are able to measure a $< 10 \%$ change in $\Delta S(T)$.

**DISCUSSION**

Structural DNA nanotechnology’s recent success can be attributed to the unusual ease with which even novices can design, fabricate, and image new nanostructures. However, imaging resolution limits the field to purely qualitative measurements of yield. Bridging the gap to implementation demands quantitative measures of the yield for all necessary, active components. If a plasmonic structure fails to contain functional strands addressing metallic nanoparticles to their correct locations, or a drug delivery system is missing a targeting moiety, it will be of little solace that the structure is otherwise of the correct shape. It is critical to understand the folding process sufficiently to predict such functional yields. An essential first step in predicting functional yield is to understand the driving thermodynamics.

We have performed an initial exploration of conformational entropy and bending enthalpy in a model DNA origami fold system. Our results indicate that, while existing theoretical models for loop energy, when modified for the composite construct stiffness, show a similar trend and magnitude to the observed folds, much more sophisticated modeling is needed to predict multi-step folding in full origami systems. However, our results suggest that folding proceeds along the following lines. Among the sub-sequences on a staple, the longest binds first in a bimolecular reaction, consistent with the nucleation seed subsequence concept(63). After this first event, the subsequent subsequences bind unimolecularly, sequentially, as controlled by loop entropy (Fig. 4), strongly favoring the short folds. These short folds further reduce the loop entropy for nearby staples and the system assembles progressively. As loop entropy and other cooperative energetics are sequence-agnostic, and capable of overwhelming the sequence-dependent energetics, this would explain why
qualitative experimental work has found origami folding to proceed like a traditional nucleation and growth system with predetermined nucleation sites (19, 21, 22).

We obtained these results by applying our newly-developed tools, which allow high-throughput, quantitative thermodynamic measurement of the smallest unit process of DNA origami folding. These tools enable exploration of the extensive parameter space associated with origami design and the many coupled reversible reactions which define the folding process for each possible design. The approach presented here may also be applied to the measurement of relatively small energy changes arising in biological DNA systems. While this experiment, run on basic qPCR equipment, did not resolve the even smaller differences in $\Delta H$, $\Delta S$, and $\Delta C_p$ associated with small changes in persistence length, these changes were clearly visible in melt temperature trends. Further reductions in the uncertainties of the raw data should enable the extraction of these small thermodynamic differences. In the future, we will conduct a sensitivity analysis to determine the quality and number of replicates required to extract entropic and enthalpic penalties with sufficient precision to simultaneously measure and distinguish persistence length effects and loop distance effects.

Our ability to make these measurements relies on careful design of the fluorescence reporter, sample stoichiometry, and the application of affine transformations. The latter, by removing systematic errors associated with typical fluorescence baseline corrections, enables the use of replicate data sets to significantly reduce measurement uncertainty. We are thus able to measure changes in $\Delta S(T)$ well below 10% of the total value, all with cost-effective qPCR equipment and reagents. **Importantly, the potential to extract high-quality thermodynamic data that we demonstrate here is a positive indication that, by using the entire dataset obtained in standard fluorescence DNA melting experiments useful additional information can be extracted with minimal extra effort.**

By improving *operando* characterization of origami folding, we hope to complement existing approaches to measurement of functional yield in order to develop design heuristics for DNA nanotechnology as a whole. Finally, we note that the type of variable persistence-length system we introduce here could be valuable as a model to explore the control of stiffness and topology in polymer physics. Supplementary Data are available at NAR online.

**ACKNOWLEDGEMENT**

Special thanks to the University of Maryland/NIST Institute for Bioscience & Biotechnology Research (IBBR) for access to their DSC equipment.

**REFERENCES**


TABLE AND FIGURES LEGENDS

Table 1. Melt experiment concentrations

Table 2. DSC, nearest-neighbour, and fluorescence melt temperatures

Figure 1. a) folding transition schematics for bimolecular hybridization-only (H.O.), i.e., staple binds to scaffold, and unimolecular, i.e. bound staple hybridizes fully, fold reactions. b) Persistence length strands: name, anticipated L_p, and schematic of oligomers/scaffold, c) Reporter scheme: cartoon of fold and reporter strands on the scaffold and description of sub-sequences for fold and reporter strands where X' indicates the complement to X. d) Relative positions of melt and anchor sub-sequence complements on M13 scaffold.

Figure 2. Hybridization-only (H.O.) control for the short Lp scaffold before a), and after b), affine transformation. The empty plate background, B(T), and fluorescence temperature dependence baseline, R(T), are subtracted from a to b, but are separately measured in control experiments.

Figure 3. DSC scans of reporter analog: a) only fold and scaffold strands. b) fold, scaffold, and fluorophore strands. c) all reporter strands. Red curve, derivative of fluorescence for the same sample. Peaks and sub-sequences labeled to match

Figure 4. Averaged melt data and van’t Hoff plots from the short Lp scaffold. H.O. indicates the hybridization-only control. The numbers in the legend refer to the number of bases between the melt and anchor subsequences for that curve. Error bars are one standard deviation above and below the mean of the replicate values.

Figure 5. T_m as a function of fold distance (log2) and system Lp. H.O. indicates the hybridization-only control. The single error bar represents the 0.3 °C temperature uncertainty as supplied by the manufacturer. The scatter in loop distance was added to the plot to allow the individual Tm values to be visualized.
Figure 6. van’t Hoff analysis and residuals for the short Lp, 286 base fold. The top fold data points are individual measurements; the uncertainty is captured in the spread of these points.

Figure 7. Fitted $\Delta H^{44}$ °C, $\Delta S^{44}$ °C and $\Delta C_p$ for the short Lp scaffold. Dotted lines indicate the minimal literature models for the energetic effects of folding a circular worm-like-chain. Error bars are one standard deviation above and below the mean of the replicate values. H.O. indicates the hybridization-only control, and is plotted at a loop distance of 0.
Supplementary Information:

Contents
Section 1: Anticipated Secondary Structure ........................................................................... 27
Section 2: Affine Transformation Algorithm ......................................................................... 28
Section 3: Entropy model for folding hybrid systems, estimation of persistence length, & Lₚ Strands30
Section 4: DSC data & uncertainty .......................................................................................... 31
Section 5: Hybridization-Only system comparison ................................................................. 35
Section 6: Pipetting variation & Hybridization-Only systems ............................................. 35
Section 7: ΔH₀ vs. ΔHᵣ₀(T) ..................................................................................................... 36
Section 8: Data from All Lₚ Systems ....................................................................................... 37
Section 9: Entropy Enthalpy Compensation ......................................................................... 40
Section 10: uncertainty in van’t Hoff analysis of optical melt curves ................................ 41
Section 11: fold strand sequences ......................................................................................... 41
Section 12: Extracted thermodynamics with and without low temperature monotonicity constraint42
Section 13: Discussion of heat capacity, enthalpy and entropy conventions ......................... 43
Section 14: Lₚ strand sequences ......................................................................................... 44
Section 1: Anticipated Secondary Structure

Given the unexpected pattern of melt temperatures for the native scaffold, a preliminary examination of the M13MP18 scaffold secondary structure was performed using ViennaRNA\textsuperscript{1} using parameters derived by Turner and Mathews\textsuperscript{2}. This identified a countable number of regions where the scaffold could form >9 base secondary structure.

Figure S1 shows a map of this secondary structure, where the black circle represents the M13, blue arcs indicate a subset of the engineered fold distances, while yellow, orange and red arcs indicate 9, 11, and 12 base regions of possible long-distance secondary structure.

We identified regions on the ‘left’ side of the M13 (bases 3,500 to 7,200) with the 11 or 12 base possible matches. We then identified a subset of the repeating 37 base oligos to block these regions from forming secondary structure.

As such, the blocking oligomers consisted of the 37 base oligomers from the A-side, which refers to the left side of the scaffold, bases 3,407 to 7,240. See section 15 for more details. The blocked positions shown above in Fig. S1 corresponds to positions E5-F5, G7-H7 in the A-side plate.
Section 2: Affine Transformation Algorithm:

An alternative to equation 1, for systems with significantly different fluorescence efficiency temperature dependence in the quenched and unquenched state might be:

\[ m_i(T) = \tau_{fi} R(T) \sigma(T) + \tau_{bi} B(T) + \tau_{ci} \]

Where:
- \( m_i(T) \) is the measured signal
- \( R(T) \) is the universal fluorescence efficiency
- \( B(T) \) is the background signal
- \( \sigma(T) \) is the universal ‘true’ signal
- \( \tau_{fi}, \tau_{bi}, \tau_{ci} \) are affine parameters

**Steps:**
1. Identify form of Eq. 1
2. Design control samples for each term & acquire data
3. Solve for affine parameters \( \tau_{fi}, \tau_{bi}, \tau_{ci} \) for each replicate that collapse to a universal background w/ some relative error
4. Universal background function \( B(T) = \tau_{bi} B_i(T) + \tau_{bi} \)
5. Universal fluorescence efficiency function \( R(T) = \tau_{fi} R_i(T) + \tau_{bi} B(T) + \tau_{ci} \)
6. Universal ‘true’ signal \( \sigma(T) = \tau_{ai} \sigma_i(T) + \tau_{bi} B(T) + \tau_{ci} \)
7. Refine Eq. 1 or constraints (e.g. slope = 0 at high & low temp)
8. Examine residuals, ensure if have consistent/reasonable values
9. Complete
In this case, another subjective user decision would be whether the linear transformations for quenched and unquenched efficiency are identical or not. Given that both are linked to pipetting variation on fluorophore content, this might be a reasonable choice.

Another alternative to equation 1, for systems with intercalating dyes might be:

\[
m_i(T) = \tau_{fi} R_i(T) \sigma(T) + \tau_{qi} Q_i(T) [1 - \sigma(T)] + \tau_{bl} B(T) + \tau_{cl}
\]
Section 3: Entropy model for folding hybrid systems, estimation of persistence length, & $L_p$ Strands

In order to remove the potential effects of secondary structure, as well as to determine if differences in scaffold persistence length, $L_p$, resulted in measurable differences in melt temperature, $T_m$, the persistence length of the scaffold was varied.

The scaffold, excepting the melt, anchor, and quencher subsequences, was approximately 7,151 bases in length. This in turn was divided into $\approx 190$ sections, each 37 bases in length. For each section, a 37 base oligo complement and a 32 base oligo complement were ordered, SI Section 15. These oligos were selectively added to the samples to force the scaffold to have repeating lengths of ssDNA/dsDNA, the content of which would change the apparent $L_p$ of the polymer as a whole. It is worth emphasizing that the apparent $L_p$ is only a simple placeholder for the exact behavior of such a complex system.

While the $L_p$ systems afforded us both freedom from secondary structure as well as control over the polymer flexibility, it significantly complicated application of any model to interpret the trends in excess entropy and enthalpy.

The Jacobson-Stockmayer entropy extrapolation, described in the top of Figure S2, models the entropy change from a line/loop. This is used to model a loop to fold transition as described in the middle/bottom of Figure S2.

It is critical to note that the units for $N$ are in bases of ssDNA. However, our hybrid systems have both ssDNA and dsDNA.

To address this, we needed to create an estimated conversion factor between the units ‘bases ssDNA’ and ‘bases of hybrid system.’

To resolve this difference, we made the aggressive assumption that the alternating ssDNA and dsDNA subsections were so small (37 bases) compared to the scaffold that they behaved homogenously.

**Figure S9: Entropy model for folding**
If this was true, the number of Kuhn lengths in a single repeat of the hybrid would equal the number of Kuhn lengths of ssDNA in a single repeat plus the number of Kuhn lengths of dsDNA in that repeat. As such, within the repeat unit of the composite construct we could take the number of bases of the dsDNA segment and multiply it by the ratio of the Kuhn lengths, or persistence lengths, of the two polymers to determine the number of equivalent ssDNA bases that segment of dsDNA contributed. This is then added to the number of bases of ssDNA in the repeat unit to find the equivalent number of ssDNA bases in each repeat. This is then divided by the total number of bases in the composite construct to make a conversion factor between the units “bases ssDNA” and “bases composite construct” as in Eq. S1.

$$\frac{\text{bases of ssDNA}}{\text{base of hybrid repeat}} = \frac{\text{bases}_{\text{ssDNA-per-repeat}} \times \text{bases}_{\text{dsDNA-repeat}}}{\text{#bases per repeat}}$$

For this, the persistence length of dsDNA in Mg buffer was taken to be 50 nm, and the persistence length of ssDNA was taken to be 3 nm.³ It is worth explicitly stating that this is obviously a suboptimal modeling approach. However, as there is no simple and tested model for enthalpies of bending and configurational entropies of systems with repeating changes in persistence length, and as the purpose of this study is measurement of small changes in energy due to topological changes, a precise and thorough theoretical framework is well beyond the scope of this work.

<table>
<thead>
<tr>
<th>System</th>
<th>Bases ssDNA per Hybrid repeat</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘Short’</td>
<td>0.168</td>
</tr>
<tr>
<td>‘Medium’</td>
<td>0.159</td>
</tr>
<tr>
<td>‘Long’</td>
<td>0.0448</td>
</tr>
</tbody>
</table>

Table S3. Conversion factors for entropy model

Enthalpy was modeled by the worm-like-chain model, assuming the energy of bending a rigid segment is entirely enthalpic in nature⁴. \( L_p \) is the persistence length, \( L_0 \) is the contour length for loop being measured (N⁵), \( R \) is the radius of curvature, and \( T \) is the temperature.

$$E_b = \frac{L_p K_b T}{2 R^2} \Rightarrow \Delta H_{\text{bend}} = \frac{L_p K_b T}{2 R^2} \times L_0 = \frac{L_p K_b T}{2 \left( \frac{L_0}{2 \pi} \right)^2} \times L_0$$

Section 4: DSC data & uncertainty

While DSC is a highly useful tool for comparing \( \Delta H \) for a reaction between samples, there were several reasons the data reported in the main text did not include analysis of DSC data.
The error associated with baseline correction in attempting to extract $\Delta H$ for comparison to van’t Hoff analysis is significant. Below are the raw, and buffer scan corrected data for the full DSC model system.

While the software can analyze the data without baseline correction, this involves 5 fit parameters per peak (offset, slope, $\Delta H$, $\Delta C_p$, $T_m$) with two of the parameters. Due to the overlap in first two terms, attempting to fit using these general models was infeasible for our reporter system.

Less general 2-state models can be fit to the data, however these require neglect of $\Delta C_p$ and baseline correction. While the former is unfortunate, the latter is prohibitive, as we will show below.

Numerous potential functions for baseline correction are available, for this illustration it was sufficient to use the most linear baselines and a 2nd order/linear baseline pair. Table S2 shows several different baseline type and window choices and their result on the calculated values.

The variation between $\Delta H$ values, combined with the inability to fit a general function (and thus obtain $\Delta C_p$) is problematic. A reporter with more distinct melt peaks would likely have been much more tractable to analysis.
<table>
<thead>
<tr>
<th>α</th>
<th>B</th>
<th>δ</th>
</tr>
</thead>
<tbody>
<tr>
<td>ΔH (kCal/mol)</td>
<td>38.63</td>
<td>143.3</td>
</tr>
<tr>
<td>T_m (°C)</td>
<td>52.55</td>
<td>81.27</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>α</th>
<th>B</th>
<th>δ</th>
</tr>
</thead>
<tbody>
<tr>
<td>ΔH (kCal/mol)</td>
<td>53.99</td>
<td>143.1</td>
</tr>
<tr>
<td>T_m (°C)</td>
<td>45.78</td>
<td>81.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>α</th>
<th>B</th>
<th>δ</th>
</tr>
</thead>
<tbody>
<tr>
<td>ΔH (kCal/mol)</td>
<td>38.63</td>
<td>143.3</td>
</tr>
<tr>
<td>T_m (°C)</td>
<td>52.55</td>
<td>81.27</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>α</th>
<th>B</th>
<th>δ</th>
</tr>
</thead>
<tbody>
<tr>
<td>ΔH (kCal/mol)</td>
<td>53.99</td>
<td>143.1</td>
</tr>
<tr>
<td>T_m (°C)</td>
<td>45.78</td>
<td>81.1</td>
</tr>
<tr>
<td></td>
<td>$\Delta H$ (kCal/mol)</td>
<td></td>
</tr>
<tr>
<td>----------------------</td>
<td>------------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td></td>
<td>44.36</td>
<td>136.9</td>
</tr>
<tr>
<td></td>
<td>50.91</td>
<td>81.44</td>
</tr>
</tbody>
</table>

Table S4: Examples of DSC data analysis
Section 5: Hybridization-Only system comparison

Below are the fitted $\Delta H(44^\circ C)/\Delta S(44^\circ C)$ values for the **Hybridization-Only** controls in all $L_p$ systems. As one would expect from the $T_m$ values reported in the main text, the $\Delta H/\Delta S$ values differ between the native/blocked and short/medium/long $L_p$ samples. The DSC control is different in $\Delta H/\Delta S$ although its $T_m$ as reported in the main text, is consistent.

The fluorescence measurements of the DSC system required measuring a 0.75 uL droplet in oil to simultaneously maintain constant concentration and avoid overexposing the qPCR optical detector. Another potential source of discrepancy is the high magnesium concentration used in DNA nanotechnology, for which the corrections to the nearest neighbor model may be insufficient. As such it is difficult to definitively say why the van’t Hoff $\Delta H/\Delta S$ values are different between these systems.

<table>
<thead>
<tr>
<th></th>
<th>$dH$ (kCal/mol)</th>
<th>error</th>
<th>$dS$ (cal/mol·K)</th>
<th>error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short</td>
<td>69.7</td>
<td>0.9</td>
<td>193.4</td>
<td>2.9</td>
</tr>
<tr>
<td>Medium</td>
<td>72.3</td>
<td>1.2</td>
<td>201.7</td>
<td>3.9</td>
</tr>
<tr>
<td>Long</td>
<td>72.4</td>
<td>4.4</td>
<td>202.0</td>
<td>14.0</td>
</tr>
<tr>
<td>Unblocked</td>
<td>57.5</td>
<td>0.9</td>
<td>155.5</td>
<td>2.9</td>
</tr>
<tr>
<td>Blocked</td>
<td>55.2</td>
<td>1.7</td>
<td>147.9</td>
<td>5.3</td>
</tr>
<tr>
<td>DSC</td>
<td>103.7</td>
<td>1.6</td>
<td>298.3</td>
<td>4.9</td>
</tr>
</tbody>
</table>

Section 6: Pipetting variation & Hybridization-Only systems

When $\Delta H$, $\Delta S$, & $\Delta C_p$ are fitted independently the $\Delta S$ values for the no fold system are higher than anticipated, as discussed in the main text. One possible explanation for this is that the no-fold system is bimolecular more susceptible to variation in the concentration (or pipetting variation). While our experimental system was designed to minimize the effect of pipetting variation, initial error in concentration measurements is possible.

To examine the possible magnitude of this effect, the analysis was rerun 3 times assuming different pipetting variation on the scaffold and fold strands.

<table>
<thead>
<tr>
<th>Percent Variation</th>
<th>Concentration</th>
<th>$\Delta H$</th>
<th>$\Delta S$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scaff  Fold</td>
<td>Scaff Fold</td>
<td>$\Delta H$</td>
<td>$\Delta S$</td>
</tr>
<tr>
<td>0  0</td>
<td>45 225</td>
<td>70.0 ± 1.1 KCal/mol</td>
<td>195.0 ± 3.5 cal/mol-K</td>
</tr>
<tr>
<td>-5%  5%</td>
<td>42.75 236.25</td>
<td>69.7 ± 1.1 KCal/mol</td>
<td>194.2 ± 3.5 cal/mol-K</td>
</tr>
<tr>
<td>5%  -5%</td>
<td>47.25 213.75</td>
<td>70.0 ± 1.1 KCal/mol</td>
<td>196.0 ± 3.5 cal/mol-K</td>
</tr>
</tbody>
</table>
While 5% error is more than would be anticipated for these systems, even 10% error in opposite directions for the scaffold and fold strands result in < 2 cal/mol-K variation in ΔS. As such it is highly unlikely that the pipetting variation meaningfully contributes to differences in the extracted ΔS.

Section 7: ΔH° vs. ΔH°(T)
As discussed in the main text, one can assume that ΔH is temperature independent, ΔH°, or assume that it follows Sup. Eq. s2. As such, it is worth comparing ΔH(T) to ΔH° as a measure of whether fitting the curvature in van’t Hoff analysis is overfitting noise.

\[ ΔH(T) = ΔH° + ΔC_p ΔT = ΔH° + ΔC_p (T_{eval} - T_m) \]  

Whether or not temperature independence was assumed, the same values for ΔH° were calculated, within error. Fig S4. shows an example plot comparing the two quantities.

If the ΔH° values from both fits considering and neglecting ΔC_p diverged, it might imply that the system was over defined, and the fitting algorithm was using play between ΔH° and ΔC_p to overfit noise in the data.

Phrased another way, this indicates that fit of the slope of the van’t Hoff plot is the same whether or not curvature is accounted for.
Section 8: Data from All $L_p$ Systems

Subsection 8a: $\Delta H(T)/\Delta S(T)$ from all $L_p$ systems

Below are the $\Delta H$ and/or $\Delta S$ for the various persistence length systems. Overall, the native/blocked systems to have too much variation in secondary structure to provide useful information. It is worth mentioning that some of these systems failed to achieve 5% relative error between curves during affine transformation. Similarly, while it is possible to distinguish the small changes in $\Delta S$ as a function of loop distance, when this is combined with simultaneous small changes in $\Delta H$ associated with increased persistence length at short distances it becomes difficult to identify a trend in the data. It is possible that with higher quality data the differences in $\Delta H$ and/or $\Delta S$ under these conditions could be resolved.
Subsection 8b: Average fluorescence from all systems

The table below shows the average melt traces for all L_p systems.

<table>
<thead>
<tr>
<th>Native scaffold</th>
<th>Blocked scaffold</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Native scaffold graph" /></td>
<td><img src="image2" alt="Blocked scaffold graph" /></td>
</tr>
<tr>
<td><img src="image3" alt="Short Lp graph" /></td>
<td><img src="image4" alt="Medium Lp graph" /></td>
</tr>
<tr>
<td><img src="image5" alt="Long Lp graph" /></td>
<td></td>
</tr>
</tbody>
</table>

The table shows the average melt traces for all L_p systems.
Section 9: Entropy Enthalpy Compensation

For literature discussion of entropy/enthalpy compensation, we refer the reader to the following review. While this compensation can be real, it is often an artifact resulting from measurement uncertainty or neglect of $\Delta C_p$.\(^5\)

The figures below show plotted entropy and enthalpy values for the linear fit, neglecting $\Delta C_p$, on the right, with the same plot is shown on the left for values fitted including $\Delta C_p$. It is worth noting that the error bars in these figures and in the main text are based on the standard deviation between replicates and don’t include the quality of fit. While the error bars in the figure on the left are larger, the residuals on this fit are much lower. It is more accurate to consider the error bars on the left as an underrepresentation of uncertainty.

The important distinction between these figures is the extent to which the values in the linear fit are spread out evenly across compensating $\Delta H$ and $\Delta S$, while the extracted values are much more clustered when $\Delta C_p$ is fitted. This indicates that much of the observed $\Delta H-\Delta S$ compensation is statistical in nature.
Section 10: uncertainty in van’t Hoff analysis of optical melt curves

The propagation of uncertainty in van’t Hoff analysis is worth consideration if one wishes to extract thermodynamic parameters from optical melt curves. Assume that the uncertainty on the true signal is $\pm N$, as below.

$$\sigma(T) = \sigma(T) \pm N$$

Given that van’t Hoff analysis is performed on $\ln(K)$ vs $1/T$, one calculates the equilibrium constant, shown below for unimolecular reactions.

$$K(T) \pm \delta = \frac{[\text{ssDNA}]}{[1 - \text{ssDNA}]} = \frac{[\sigma(T) \pm N]}{[1 - \sigma(T) \pm N]}$$

$$\delta = \sqrt{\left(\frac{N}{\sigma(T)}\right)^2 + \left(\frac{N}{1 - \sigma(T)}\right)^2}$$

As the melt curve, $\sigma(T)$, approaches either plateau, i.e., as $[\text{ssDNA}]$ approaches 0, or $[1 - \text{ssDNA}]$ approaches 0, the equilibrium constant is dominated by noise. As such, in fitting $\ln(K)$ vs $1/T$, we only used the middle 9 points, or 5.5 °C, of the melt curve.

Section 11: fold strand sequences

The subsequences listed below match those described ($\alpha, \beta', \beta, \gamma, \delta$) in Fig. 1c of the main text. As such the full strand sequences contain combinations of these with different anchor sequences, $\gamma$, to enforce different fold distances.

<table>
<thead>
<tr>
<th>Floresrophore (δ)</th>
<th>GTC CAC TAT TAA ACG TGG TGG CCA ACG TCA ACG TCA TCA AAG GGC GAA - JOE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quench Complement (β')</td>
<td>C GCT GAT AAC TGT CCG GTA TAA CGT CCG TAG CAG GCG</td>
</tr>
<tr>
<td>Quench (β)</td>
<td>Iowa Black Quencher 5' - CGC GTC GTA CCG ACC CTA TAC CCG ACA CAC CAG GAC G</td>
</tr>
<tr>
<td>Melt (α)</td>
<td>AAA CCG TCT A</td>
</tr>
<tr>
<td>Translated Scaffold</td>
<td>GGCACCGTATGCGGATGCGTATGCGCTATAGCCCTTGGACGCCCATCTTCTTTTAGATGGAGTCGCCACGGTCCTT</td>
</tr>
<tr>
<td>Hybridization Only</td>
<td>CGC TGA TGG TAA CTA TGC GTA GGT ATA ACG TGG GTA GCA GCC GAA ACC GTC TA</td>
</tr>
<tr>
<td>Fold - 138</td>
<td>CCA CGC TGG TTT GGG CCA GCA GCC GAA AAC GCT GAT GGT AAG TCT GGT CCG GTA TAA CTT CGT CAG GCG CAG GGG CCA AAG CCG TCT A</td>
</tr>
<tr>
<td>Fold - 212</td>
<td>TCT TTT CAC CAG CGA GAC GGG CAA CAG CTC GCT GAT GGT AAG TCT GGT CCG GTA TAA CTT CGT CAG GCG CAG GGG CCA AAG CCG TCT A</td>
</tr>
<tr>
<td>Fold - 286</td>
<td>CGT GCC AGC TGC ATT AAT GAA TGG GCC AAC GCT GAT GGT AAG TCT GGT CCG GTA TAA CTT CGT CAG GCG CAG GGG CCA AAG CCG TCT A</td>
</tr>
<tr>
<td>Fold - 434</td>
<td>TTT CCT GTG TGA ATT TGT TAT CCT GTC ACC GCT GAT GGT AAG TCT GGT CCG GTA TAA CTT CGT CAG GCG CAG GGG CCA AAG CCG TCT A</td>
</tr>
<tr>
<td>Fold - 804</td>
<td>ACG TTT GTG TAG ATG GCC TGG TCA TAA CCC GCT GAT GGT AAG TCT GGT CCG GTA TAA CTT CGT CAG GCG CAG GGG CCA AAG CCG TCT A</td>
</tr>
</tbody>
</table>
Section 12: Extracted thermodynamics with and without low temperature monotonicity constraint

Without monotonicity constraint

With monotonicity constraint

The monotonicity constraint primarily affected short fold distances as shown in the figure above. There were only approximately 6 data points below the melt curve to which this constraint could reasonably be applied. As such, it is likely that the improvement in analysis quality when removing the constraint are due to insufficient data points for its proper application, rather than some fundamental failure of the constraint.
Section 13: Discussion of heat capacity, enthalpy and entropy conventions

To be brief, we here use conventions for entropy and enthalpy already existing in the literature\textsuperscript{5-7} with the understanding that they may or may not continue to be the primary convention.

By this definition $\Delta H(T)$ and $\Delta S(T)$ are the slope and intercept of the derivative of the $\ln(K)$ vs $1/T$ curve. As such, both $\Delta H(T)$ and $\Delta S(T)$ vary as a function of temperature, where the temperature dependence of both (as well as the curvature of the plot) are determined by the $\Delta C_p$.

The benefit of this convention is that it describes the change in entropy and enthalpy of a transition, at a given temperature, while accounting for the difference in heat capacity between the two states.
Section 14: $L_p$ strand sequences

As a brief description, the strands used to control the persistence length, $L_p$, described in Fig. 1 of the main text remove the secondary structure and help to create a consistent $L_p$ for the scaffold being folded.

For pipetting, the strands were divided between the “A-Side” and “B-Side.” The “A-Side” strands consist of those hybridizing to the scaffold from base 3,500 to 7,429 and were separately named as no folds occurred on that side. The “B-Side” strands hybridized from base 0 to 3,500 which was the same range over which the folds occurred. A master stock of “B-Sides” was created, described below, omitting those 24 strands which could compete with any of the 12 fold positions. For each individual fold system, an additional stock of 22 “B-FoldSpecific” strands was created to fill in the 11 positions not being folded.

So, for every system, the same master “A-Side” and master “B-side” stock were added, as well as a different “B-FoldSpecific” stock. This was done to minimize potential variation between fold systems to as few strands as possible.

The blocking strands (section 1) consisted of A-Side, 37 bp oligos from wells E5-F5 and G7-H7.

| A-Side       | The half of the scaffold not being used for folding |
| B-Side       | The half of the scaffold with folding |
|              | On the B-Side pools, the fold sections must be left out |

A-Side Pools

| Short | All strands from A-Side_32bpOligos |
| Long  | All strands from A-Side_37bpOligos |
| Medium | 32 bp wells | A1,A3,A5,...A11,B1,B3,...H11 |
|        | 37 bp well   | A2,A4,A6,...A12,B2,B4,...H12 |

B-Side master stocks

The 24 persistence length strands which *could* compete with any arbitrary fold distance were left out of the master stocks. Specific stocks for each fold distance, omitting the 2 strands which would compete with that fold were pipetted separately.

Strands Left out from master stocks

| omitted | A1,A6, A9, A10, B7, B8, C5, C6, D3, D4, E1, E2, E11, E12, F9, F10, G7, G8, H5, H6 |
| Short   | All strands not omitted from B-Side_32bpOligos |
| Long    | All strands not omitted from B-Side_37bpOligos |
| Medium  | 32 bp wells | A7,A11,B1,B3,...H11 |
|         | 37 bp well   | A8,A12,B2,B4,...H12 |

Below are which folds compete with which well positions on the B-Side...
Position from B-Side which would block the anchor for that fold.

<table>
<thead>
<tr>
<th>Fold</th>
<th>A1, A2</th>
<th>A3, A4</th>
<th>A5, A6</th>
<th>A9, A10</th>
<th>B7, B8</th>
<th>C5, C6</th>
<th>D3, D4</th>
<th>E1, E2</th>
<th>E11, E12</th>
<th>F9, F10</th>
<th>G7, G8</th>
<th>H5, H6</th>
</tr>
</thead>
<tbody>
<tr>
<td>138</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>212</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>286</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>434</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>804</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1174</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1544</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1914</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2284</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2655</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3025</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3394</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

DNA Sequences

### A-Side, 32bp oligos

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Pos</th>
</tr>
</thead>
<tbody>
<tr>
<td>tccagtaagcgtcatacatggttttgatg</td>
<td>3690</td>
</tr>
<tr>
<td>aagccagaatggaaagcgcagtctctgaat</td>
<td>3727</td>
</tr>
<tr>
<td>ttggcccttgatattcacaacaacgaatggatctt</td>
<td>3764</td>
</tr>
<tr>
<td>cgccacgacctgagaggagtgtgagccagtca</td>
<td>3801</td>
</tr>
<tr>
<td>ccctcagagccgccaccagaacccccaccacca</td>
<td>3838</td>
</tr>
<tr>
<td>gcgcgccaccctcagaaacc gcaccaccacca</td>
<td>3875</td>
</tr>
<tr>
<td>cggaaccagagcaccaccgccgccacgctccccc</td>
<td>3912</td>
</tr>
<tr>
<td>ttattacgcgttgcctttcttccataacaa</td>
<td>3949</td>
</tr>
<tr>
<td>tagcgcttttttcataagcagctttaggcctatag</td>
<td>3986</td>
</tr>
<tr>
<td>tagcgcacagaataagtttgctcttagcgtca</td>
<td>4023</td>
</tr>
<tr>
<td>tcaccaagaaaacctcagtagacgccgcta</td>
<td>4060</td>
</tr>
<tr>
<td>accagtgccaccattaccattgcaaggcccgg</td>
<td>4097</td>
</tr>
<tr>
<td>cgacctgagccatttgggaatagagccgacaca</td>
<td>4134</td>
</tr>
<tr>
<td>cggaattaattcattaaaggtgaattatcacc</td>
<td>4171</td>
</tr>
</tbody>
</table>

### A-Side, 37bp oligos

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Pos</th>
</tr>
</thead>
<tbody>
<tr>
<td>tccagtaagcgtcatacatggttttgatgat</td>
<td>3690</td>
</tr>
<tr>
<td>aagccagaatggaaagcgcagtctctgaatttacg</td>
<td>3727</td>
</tr>
<tr>
<td>ttggcccttgatattcacaacaacgaatggatcttatta</td>
<td>3764</td>
</tr>
<tr>
<td>cgccacgacctgagaggagtgtgagccagtca</td>
<td>3801</td>
</tr>
<tr>
<td>ccctcagagccgccaccagaacccccaccacca</td>
<td>3838</td>
</tr>
<tr>
<td>gcgcgccaccctcagaaacc gcaccaccacca</td>
<td>3875</td>
</tr>
<tr>
<td>cggaaccagagcaccaccgccgccacgctccccc</td>
<td>3912</td>
</tr>
<tr>
<td>ttattacgcgttgcctttcttccataacaaatcag</td>
<td>3949</td>
</tr>
<tr>
<td>tagcgcttttttcataagcagctttaggcctatagccc</td>
<td>3986</td>
</tr>
<tr>
<td>tagcgcacagaataagtttgctcttagcgtca</td>
<td>4023</td>
</tr>
<tr>
<td>tcaccaagaaaacctcagtagacgccgcta</td>
<td>4060</td>
</tr>
<tr>
<td>accagtgccaccattaccattgcaaggcccggaa</td>
<td>4097</td>
</tr>
<tr>
<td>cgacctgagccatttgggaatagagccgacaaatc</td>
<td>4134</td>
</tr>
<tr>
<td>cggaattaattcattaaaggtgaattatcaccgta</td>
<td>4171</td>
</tr>
</tbody>
</table>
gccagtaataagagaatataaagtaccgacaaa
5207
caacgcacaacatgtaatttaggcagaggcatt
5244
tcaacagttagggccttaattgagaatcgcctacaata
5281
tgcgttatacaaaattttcaccgtaaagaacgc
5318
aatcataattactagaagaaagccttgttagta
5355
ttgataaaaatagggcttaaataaagaataaaacaccgg
5392
cttctgacccaatttaattggtttaaatacc
5429
agaaaaaccctttccaatatatttttagtaat
5466
aatgctgtagcacaatcacaatgcgaagacaagac
5503
aacctccggcttaggggttatataactatatga
5540
atttatcaaaatcataggtctgagactactcctttt
5577
tagttgagattaaggctgagagaagtcataatgta
5614
aatatatattccttagaaatctcttgaacacat
5651
tgatgtaaataacctgtctgtgtaaatctgtcctatt
5688
ccttttttaatggaacagtacaataaataat
5725
acaaaatattacacatttaaactttcatgatga
5762
cagagcaaaaaagagatgaagaaacaacatca
5799
caaaatccgccccagggcaattatcttttca
5836
ataacggattcgcctgtggttgtttgaataccagaatcgttaa
5873
atatacagtaacagtccttttacatcgggag
5910
agaaatgtctgtagattttcaggttttaacgctca
5947
accatatcaaatattagcagcataaaccaga
5984
tggattatacttctgttaatttagggaaggaatttg
6021
ccagaagagcgcaggtatatcataatttcctgtaa
6058
ccagaagagcgcaggtatatcataatttcctgattat
6095
atttagtaacattatcatttttcggaacacca
6132
ttgaatcctttgccccaggttattaattttta
6169
gaagtattagactttacaacaatgcaaac  
attagagccgctaatagataatacatatgagg  
tatctaaatatcttttagttgcaactaacaact  
gttgccaatacacaactgtagaagaatgtggagg  
aaccctaaatatcaaccctctaataatcatctct  
cagcagaaatgaaataatcaagcatcaacct  
gtattaacaccgcctgcaacagtgcacgcgctg  
accaccagcagaagataaaacagaggtgaggtgggc  
atttttgaatggctattagtctttaatgcgcg  
ctgacctgaaagcgtaagaatacgtggcacag  
taaaagggacattctggccaacagagatagaa  
tttacattggcagattcaccagtcacacgacc  
cacgcaaattaacgcttgtagcaatacttctt  
taatcagtgaggccacggcttagtaagaagactgctg  
caggaacggtacgccagaatcttgagaagtt  
gggagctaaacaggaggccgattaaagggatt  
agcacgtataaactgctttcctctgtgggaatc  
ccggcgaacgtggcgagaaaggaagggaagaa  
taaaagggagcccccgatttagagcttgacggg
tgggtcgaggtgccgtaaagcactaaatcgg

7205
tgggtcgaggtgccgtaaagcactaaatcggaaccc

7205

B-side Oligos, 32bp

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Pos</th>
</tr>
</thead>
<tbody>
<tr>
<td>gatggtgttcgaaatcggcaaaatccctta</td>
<td>143</td>
</tr>
<tr>
<td>ccacgtgtggttgccccacagggcgaataatcc</td>
<td>180</td>
</tr>
<tr>
<td>tttcaccgctgcctagagagttgacagcaaa</td>
<td>217</td>
</tr>
<tr>
<td>tcttttcaccgagacggggaacacagtcat</td>
<td>254</td>
</tr>
<tr>
<td>gagagccggtttgctattgggccagggtg</td>
<td>291</td>
</tr>
<tr>
<td>cgtgcacagtcgattgaatcgcgcaacgct</td>
<td>328</td>
</tr>
<tr>
<td>tgtgccttcactgccccgtttccagtcgga</td>
<td>365</td>
</tr>
<tr>
<td>ggggtcctaattgaatgtagtaaacatcataa</td>
<td>402</td>
</tr>
<tr>
<td>acaacatagcagccggaacgataaaattgtaa</td>
<td>439</td>
</tr>
<tr>
<td>tttctgtgtgaatttttcagtcataaatcacaat</td>
<td>476</td>
</tr>
<tr>
<td>ggtacgagctcgaattcctaggtagtcat</td>
<td>513</td>
</tr>
<tr>
<td>gttgcatgcctcgagttgcaacttaagagat</td>
<td>550</td>
</tr>
<tr>
<td>cccagtcaacgcgtgtaaactgagccgagct</td>
<td>587</td>
</tr>
<tr>
<td>gtcgcaaggttaggttgtaaccgccagg</td>
<td>624</td>
</tr>
<tr>
<td>ctcttcgttattacgcagcattggcgaagaggg</td>
<td>661</td>
</tr>
<tr>
<td>ggtgctgcaactggtggtggaagggcgttctg</td>
<td>698</td>
</tr>
<tr>
<td>gttgccggaaacgcgcaacgccttgccatcc</td>
<td>735</td>
</tr>
<tr>
<td>aagatgcaactcagcagctttccgccaccg</td>
<td>772</td>
</tr>
<tr>
<td>gccagtgggttaggggacagcgcgcttcgccc</td>
<td>809</td>
</tr>
<tr>
<td>acgtttggttagatgggccatcgaaccgctg</td>
<td>846</td>
</tr>
<tr>
<td>gggaacacagcgccggtagctgcataggatct</td>
<td>883</td>
</tr>
<tr>
<td>taaatgtagcgagtaacaacccctgctgattc</td>
<td>920</td>
</tr>
<tr>
<td>cgctctgctcctctgcagcactcatcaccat</td>
<td>957</td>
</tr>
</tbody>
</table>

B-side Oligos, 37bp

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Pos</th>
</tr>
</thead>
<tbody>
<tr>
<td>gatggtgttcgaaatcggcaaaatccctta</td>
<td>143</td>
</tr>
<tr>
<td>ccacgtgtggttgccccacagggcgaataatcc</td>
<td>180</td>
</tr>
<tr>
<td>tttcaccgctgcctagagagttgacagcaaa</td>
<td>217</td>
</tr>
<tr>
<td>tcttttcaccgagacggggaacacagtcat</td>
<td>254</td>
</tr>
<tr>
<td>gagagccggtttgctattgggccagggtg</td>
<td>291</td>
</tr>
<tr>
<td>cgtgcacagtcgattgaatcgcgcaacgct</td>
<td>328</td>
</tr>
<tr>
<td>tgtgccttcactgccccgtttccagtcgga</td>
<td>365</td>
</tr>
<tr>
<td>ggggtcctaattgaatgtagtaaacatcataa</td>
<td>402</td>
</tr>
<tr>
<td>acaacatagcagccggaacgataaaattgtaa</td>
<td>439</td>
</tr>
<tr>
<td>tttctgtgtgaatttttcagtcataaatcacaat</td>
<td>476</td>
</tr>
<tr>
<td>ggtacgagctcgaattcctaggtagtcat</td>
<td>513</td>
</tr>
<tr>
<td>gttgcatgcctcgagttgcaacttaagagat</td>
<td>550</td>
</tr>
<tr>
<td>cccagtcaacgcgtgtaaactgagccgagct</td>
<td>587</td>
</tr>
<tr>
<td>gtcgcaaggttaggttgtaaccgccagg</td>
<td>624</td>
</tr>
<tr>
<td>ctcttcgttattacgcagcattggcgaagaggg</td>
<td>661</td>
</tr>
<tr>
<td>ggtgctgcaactggtggtggaagggcgttctg</td>
<td>698</td>
</tr>
<tr>
<td>gttgccggaaacgcgcaacgccttgccatcc</td>
<td>735</td>
</tr>
<tr>
<td>aagatgcaactcagcagctttccgccaccg</td>
<td>772</td>
</tr>
<tr>
<td>gccagtgggttaggggacagcgcgcttcgccc</td>
<td>809</td>
</tr>
<tr>
<td>acgtttggttagatgggccatcgaaccgctg</td>
<td>846</td>
</tr>
<tr>
<td>gggaacacagcgccggtagctgcataggatct</td>
<td>883</td>
</tr>
<tr>
<td>taaatgtagcgagtaacaacccctgctgattc</td>
<td>920</td>
</tr>
<tr>
<td>cgctctgctcctctgcagcactcatcaccat</td>
<td>957</td>
</tr>
</tbody>
</table>
ctttaacagttcagaaaacgagaatgaccat
1993
gaatgctcataaatatcctgttaatccccctc
2030
agtaaaatgtttgactgttgatagctgcataaatcgcg
2067
aggcttttgcaaaaagaagtttttgcagaggagggg
2104
tcgtttaccagacgacgataaaaacacacaaata
2141
attacgagccatagtaagacaacactatcataac
2178
cacattcaactataacgataccataacgcaca
2215
cagtaggaagattccatgtagttgatgtgatgataac
2252
tagtaaaaacccatggtgatgtgatggatgtgatgatttca
2289
gcttactataccagctaggactgctgagaaagaagaaagtttttgcaaaagaagttttgccagagggg
2326
taagggaaccgaactataacgccggaaccgaacca
2363
taggtaaattgggcttgagatggtttaatttccatgttaaatgaccgtgtctgtgatttcaattttca
2400
tgaataaggcttgcccttgacgagaaacaccaga
2437
cattaccaaaatcaacgtgatcaaccaacctcgctca
2474
cctcatcaagtaatcttgacaagaaagcgcga
2511
atgaaacggtgtacagaccgcagcgcataggctg
2548
taagggaaccgaacctgcaaaccactttaaaagag
2585
ttgtacttagcgcgaagcgcggagcagctggtc
2622
gcttgataaattgtgtgaaatccgcgacctgctcaaca
2659
aaggcagacaaccaacatcaacggaggattttgta
2696
actaaaacactcactttgtgaccccccacgcatt
2733
aaggccaccaacctaaaacgaaagaggcagaagaag
2770
tttccatattacgggtaaaaatacgtaatgcca
2807
agaggcttttgagacaaagactttttcatgta
2844
cgaagacagcatgcggagcaggttagacacacg
2881
ttaagggcgccttttgcgggatgcgtacccctcgc
2918
ataaccgatatattcgggtgcgtgaggtctgca
2955
References


