UC Santa Barbara
UC Santa Barbara Electronic Theses and Dissertations

Title

DNA Nunchucks: Nanoinstrumentation for Single-Molecule Measurement of Structure-Dependent DNA Bending and Dynamics

Permalink

https://escholarship.org/uc/item/85n846fm

Author

Cai, Xinyue

Publication Date

2023

Peer reviewed|Thesis/dissertation

University of California
Santa Barbara

DNA Nunchucks: Nanoinstrumentation for Single-Molecule Measurement of Structure-Dependent DNA Bending and Dynamics

A dissertation submitted in partial satisfaction
of the requirements for the degree

Doctor of Philosophy

in

Physics

by

Xinyue Cai

Committee in charge:

Professor Deborah K. Fygenson, Chair
Professor Omar A. Saleh
Professor Boris I. Shraiman

December 2023

The Dissertation of Xinyue Cai is approved.

_____

Professor Omar A. Saleh

_____

Professor Boris I. Shraiman

_____

Professor Deborah K. Fygenson, Committee Chair

December 2023

DNA Nunchucks: Nanoinstrumentation for Single-Molecule Measurement of

Structure-Dependent DNA Bending and Dynamics

To my family,
For their endless love and support during this endeavor.

Acknowledgements

Abstract

DNA Nunchucks: Nanoinstrumentation for Single-Molecule Measurement of
Structure-Dependent DNA Bending and Dynamics

by

Xinyue Cai

We present the design and application of a novel nanoscale instrument, the "DNA nunchuck", which enables the dynamic measurement of bend angles in short double-stranded DNA (dsDNA) molecules. The DNA nunchuck, leveraging stiff, fluorescently-labeled DNA nanotubes, mechanically magnifies the orientations of an embedded dsDNA strand, allowing for precise angle determination through fluorescence imaging techniques combined with neural network image analysis. The validation of the DNA nunchuck technique is demonstrated by reproducing expected relationships between nunchuck angular variance and dsDNA length, corroborating its effectiveness in accurately measuring DNA bending. Further application of the DNA nunchuck uncovers the dynamic mechanical behavior of short dsDNA sequences, particularly the existence of meta-stable bent states that have eluded detection with previous methodologies. Detailed analysis of nunchucks with a 37 bp embedded strand identifies a significant bent state, or "kink", with a mean angle of $54 \pm 2°$, providing insight into the anomalous high cyclization frequency observed in short dsDNA. This kinked state accounts for a substantial proportion of the observed bending dynamics at room temperature. The combined findings from the DNA nunchuck's applications significantly advance our understanding of mechanical properties of dsDNA, contributing to a deeper understanding of its role in biological processes and its potential in nanoscale engineering. The work sets a precedent for future quantitative studies of DNA mechanics and dynamics at the single-molecule level.

Contents

Chapter 1

Introduction

1.1 DNA Bending and Existing Tools

The field of molecular biology has long been intrigued by the mechanics of DNA, with a particular focus on the bending dynamics that are fundamental to the regulation of gene expression and the maintenance of genomic stability. Interest in DNA bending was piqued in the early 1980s following observations of anomalous electrophoretic mobility in the kinetoplast DNA (kDNA) of trypanosomes, which hinted at the presence of intrinsically bent DNA regions within the standard B-DNA conformation[1]. Subsequent investigations revealed that these regions were associated with a distinctive pattern of periodically spaced adenine (A) tracts, commonly referred to as "A-tracts" — sequences of five to six adenines interspersed with four to six base pairs of guanine-cytosine (G+C) rich sequences[2]. These A-tracts have been discovered to be prevalent across a diverse range of prokaryotic and eukaryotic organisms, implicating their role in nucleosome positioning[3], as well as in the modulation of transcriptional activity[4].

Concurrently with the exploration of intrinsic DNA bending, it became apparent that DNA deformation is a routine byproduct of protein-DNA interactions, manifesting

universally across myriad DNA-centric biological phenomena[2, 5]. This realization has fostered a sustained interest in the thorough characterization of DNA bends, both intrinsic, which are determined by the nucleotide sequence itself, and induced, which arise from the binding of proteins and other molecules. Such characterizations aim to elucidate the structural, energetic, and dynamic aspects of these bends. Studies have investigated the physical mechanisms by which DNA bending occurs, the energetics underlying these conformational changes, and the dynamic nature of the bending process, which can be transient and influenced by the cellular environment[6, 7, 8].

In recent years, the field of DNA nanotechnology has experienced significant advancements, harnessing the unique base-pairing properties of DNA to construct a wide array of intricate nanostructures and devices[9]. These developments have opened the door to novel applications across diverse fields, ranging from condensed matter physics to medical diagnostics[10, 11, 12]. DNA origami techniques have also enabled the creation of various nanodevices, which feature not only complex three-dimensional structures but also dynamic elements with precise and programmable motion paths[13, 14]. Such progress promises novel applications in molecular robotics and nanoscale manufacturing, enhancing the versatility and functional scope of DNA-based technology. Given the burgeoning repertoire of DNA-based constructs, it is necessary to understand and quantify DNA bending characteristics, such as the inherent curvature induced by specific motifs in their relaxed states. This knowledge is crucial for the precise engineering of DNA nanostructures, enabling the fine-tuning of shapes and the functionality of devices built from DNA. For instance, bulged duplex motifs, which confer varying degrees of flexibility and can act as hinges in many DNA assemblies, has been thoroughly analyzed by simulation to help predicting and guiding the self-assembly of DNA into targeted configurations with desired mechanical properties[15].

Despite the critical importance of DNA bending, a substantial gap remains in the

dynamic and single-molecule resolution characterization of DNA bending, particularly for short double-stranded DNA (dsDNA) segments. Despite the array of methods available to measure DNA bending, each one presents significant limitations in capturing the dynamic, single-molecule resolution necessary for thorough understanding. Traditional techniques such as gel shift electrophoresis and cyclization efficiency tests yield only a crude estimate of DNA bending angles and are unable to distinguish between inherent flexibility and permanent bends[16, 17]. High-resolution methods like NMR, crystallography, and X-ray scattering, despite their detailed structural insights, are impractical for routine use due to their requirement for costly equipment, intensive sample preparation, and sophisticated analysis[18, 19, 20]. Direct visualization approaches, including Atomic force microscopy (AFM) and cryo-EM, although improving in accessibility, encounter biases from surface interactions and the limitations of 2D confinement that obscure the true dynamics of DNA bending[21, 22, 23, 24]. Optically detected methods such as fluorescence resonance energy transfer (FRET) can track bend dynamics over time, but they require extensive calibration and interpretation that indirectly measures bending through distance changes[25, 26, 8]. The recently developed loop-seq technique, while offering high-throughput evaluation of DNA cyclizability, lacks the dynamic and single-molecule aspects as well[27]. The absence of a reliable technique that can directly and simultaneously observe the bending dynamics of short strands of dsDNA at a single-molecule scale underscores the need for innovative approaches in this field.

1.2   Controversies Over Flexibility of Short dsDNA

The phenomenon of unusually high cyclization rates in short dsDNA was first documented by Cloutier and Widom, revealing a cyclization propensity in random DNA sequences approximately 100-fold greater than what would be anticipated from classical

theories of DNA rigidity[28]. This observation suggested that DNA's flexibility might be inherently greater at shorter lengths, challenging long-standing assumptions about the uniformity of DNA mechanical properties regardless of scale. Subsequent research has built a robust body of evidence, suggesting that as dsDNA strands become shorter, their effective flexibility increases. For instance, Wiggins *et al.* identified more prevalent spontaneous large-angle bends in short dsDNA than predicted by classical elasticity models[29]. Similarly, Vafabakhsh and Ha developed a novel assay revealing a biologically relevant, sharply bent DNA state with surprising flexibility in lengths under 150 bp, defying predictions of the worm-like chain model[30]. Zoli's work further supported these findings, demonstrating a significant decrease in the apparent persistence lengths in short dsDNA[31, 32]. These studies collectively indicate a more complex behavior of dsDNA flexibility than previously understood.

Skepticism towards the notion of DNA flexibility varying with length has emerged, with Vologodskii *et al.* defending traditional models of DNA bending and attributing discrepancies with Cloutier and Widom's findings to experimental variances[33]. They further critiqued Vafabakhsh and Ha's methods, suggesting that their observations of DNA cyclization can be explained by the flexibility of long sticky ends, which may not require tight alignment for cyclization, thus potentially inflating the observed cyclizability[34]. Meanwhile, Mazur *et al.* attributed the reported deviations at shorter scales observed in AFM experiments to experimental noise and errors in image analysis rather than genuine flexibility[35].

The controversy over DNA's anomalously high flexibility at short length scales remains unresolved, with recent studies contributing little to the advancement of understanding in this area. Researchers have continued to investigate DNA kinkability and cyclizability, routinely acknowledging this existing controversy in their discussions, but fresh insights or hypotheses have been notably absent [27, 36, 37, 38, 39]. This lack of

progress can be largely ascribed to the absence of suitable techniques for direct, single-molecular observation of dsDNA, which is crucial for producing the detailed data necessary to advance the debate.

1.3   Measuring DNA Bending Using DNA Nunchucks

Addressing the limitations of traditional techniques used to study DNA bending, we have developed a new tool that allows for dynamic, single-molecule quantification of DNA bending using commonly available laboratory instruments and straightforward analysis techniques. The cornerstone of this method is the creation of "DNA nunchucks"[40], a nanoscale construct featuring two rigid DNA nanotube "arms" that rigidly grow from a pair of cylindrical DNA origami "seeds". These seeds are bridged by a relatively short double-stranded DNA "linker", typically between 27 and 58 bp in length, which is the most flexible segment of the nunchuck structure, and functions as the pivot allowing the nunchuck to bend. Once constructed by self assembly, the nunchuck are free to roam and bend within a confined quasi-two-dimensional environment, constrained to a narrow space under 0.5 microns in height and in between PEG-coated glass surfaces. The fluorescently labeled nanotube arms are designed to serve as a dynamic readout of the linker's end-to-end orientation, which can be observed and recorded in real time using epi-fluorescence microscopy. Thus we can obtain a time-lapse fluorescence video of nunchucks from which the bending of the linker can be tracked. Subsequent analysis of the sequence of images from the video generates a time series of bend angles with good temporal (~ 0.1 sec) and angular ($\sim 5°$) resolution. In essence, the DNA nunchuck provides a robust, user-friendly platform for dynamic, real-time investigation of short dsDNA bending, and serves as a response to the previously unmet need for direct, single-molecule observation techniques in the study of DNA mechanics.
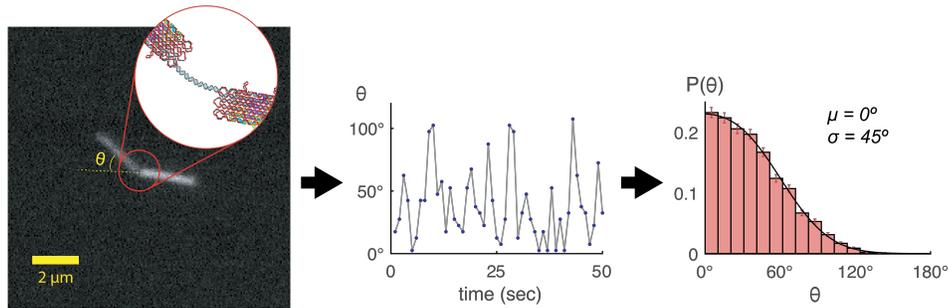
Figure 1.1: A visualization of the nunchuck concept. Left: A nunchuck captured by fluorescence video microscopy, with an inset detailing the linkage between the DNA origami seeds and the linker strand. Center: A time series representation of the bend angle fluctuations observed from a single nunchuck, demonstrating the dynamic measurement capability. Right: A histogram of bend angles from a nunchuck, which enables the extraction of the mean bend angle (μ) and standard deviation (σ) through a Gaussian fit, which are critical parameters for characterizing the bending behavior of DNA.

Subsequent sections of this dissertation detail the nunchuck's design, its assembly process, and the measurement methodologies. We then demonstrate that the variations in the linker length correspondingly alter the distribution of the nunchuck's bend angles, with changes in angular variance aligning with predictions from the worm-like chain model, the standard model for dsDNA polymer behavior. Further, we show that incorporating an intrinsically bent DNA sequence, such as phased A_6-tracts, into the linker produces a nonzero mean bend angle of nunchucks, with a magnitude that matches expectations. These results validate the nunchuck's functionality in measuring both the stiffness and inherent bending of DNA sequences, thus facilitating dynamic, single-molecule measurements of DNA bending. Beyond mere calibration for accurate DNA bending measurement, our data reveal that A_6-tract linkers exhibit a stiffness increase of $(80\pm30)\%$ compared to generic dsDNA.

Moreover, a closer examination of "outliers" among the nunchuck population, which exhibit unexpected and unique bending behavior that differs significantly from the majority of the population, reveals that the outliers are not attributable to nunchuck structural

6

defects, assembly errors, or interference with the environment. Rather, such variability in bending might represent an inherent property of the linker strand, suggesting that individual nunchucks may possess distinct bending statistics due to the presence of rare, meta-stable bent states which are not uniformly sampled within our standard observation times of ($<$20 min). This hypothesis gains support from prolonged individual nunchuck observations, which show unique bend angle distributions within different observation sub-periods. Assuming that most nunchucks explore no more than one predominant bent state within a brief observation window (~ 5 min), we employ bootstrapping methods on high-frequency imaging data (around 10 Hz) to deduce that the predominant bent state is sequence-independent, with an average bend angle of $54 \pm 2°$ and a standard deviation ranging between $21°$ and $31°$.

Chapter 2

Nunchuck Design, Assembly, and Imaging

2.1 Nunchuck Design

2.1.1 Seed

The DNA nunchuck is composed of three main parts: cylindrical DNA origami seeds, rigid DNA nanotube arms, and the double-stranded DNA linker. As introduced previously, by manipulating the length and sequence of the linker, the nunchuck's bending properties can be finely tuned, making it an ideal tool for quantifying DNA flexibility and stiffness. The seed at the heart of the nunchuck is a DNA origami structure — a meticulously arranged assembly of DNA strands folded into a predetermined shape. The ingenuity of the DNA origami seed lies in its ability to dictate the assembly and structural characteristics of the resulting nanotubes, a concept further explored in the work of Schulman *et al.* [41]. In our nunchuck design (Figure 2.1), first demonstrated by Mohammed and colleagues [42], the seed takes the form of a hollow cylinder with he-

lices that provide the nucleation sites for the growth of the nanotube arms. This design leverages the precision of DNA origami, first invented in 2006 By P. Rothemund [43], allowing for exquisite control over the resulting nanostructure's geometry.
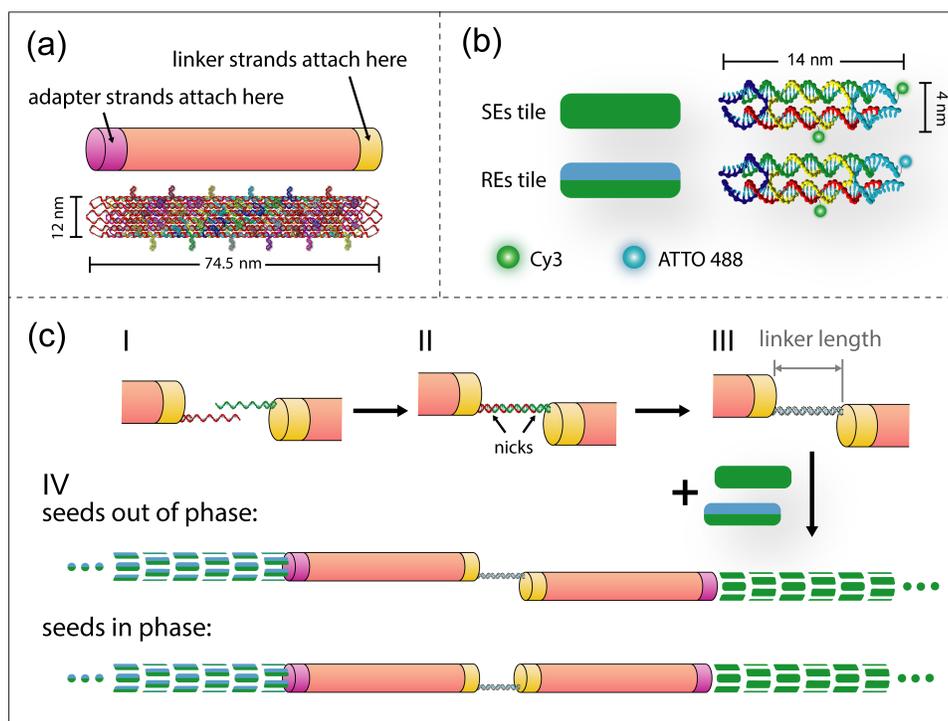

Figure 2.1: Nunchuck design and assembly. (a) Cylindrical DNA origami seeds, modified from the protocol by Mohammed *et al.*[42], are folded from a scaffold strand that is precisely 3,024 bases in length. Scaffold fringes provide specific sites for the attachment of a linker strand at one end (highlighted in yellow) and for the addition of adapter tiles at the opposite end (indicated in magenta). (b) Two distinct and fully orthogonal sets of DAE-E tiles[44] are used to construct nanotube arms that can be easily differentiated using fluorescence microscopy. Each tile of the SEs type is marked with two Cy3 fluorophores, while each tile of the REs type is labeled with one Cy3 and one ATTO 488 fluorophore. (c) Formation of seed dimers occurs as linker strands hybridize (from I to II). Following hybridization, the structures are ligated to seal the nicks and then purified by agarose gel electrophoresis (from II to III). The purified, ligated seed dimers are then introduced to a blend of SEs and REs tiles (at stage III), which initiate growth on their respective adapter strands, culminating in the self-assembly of distinctly fluorescent nanotubes. This allows for the identification and precise localization of the nunchuck's linker through fluorescence microscopy. The spatial orientation of the seeds relative to the axis of the linker is modulated by the length of the linker, conforming to the natural twist of the DNA double helix (as shown in IV).
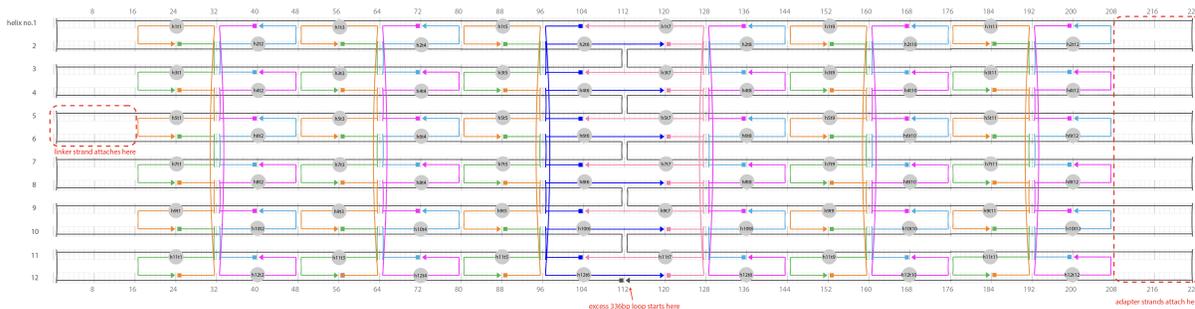
Figure 2.2: The DNA origami design of the nunchuck seed, adapted from Mohammed *et al.*[42]. Gray circles mark the positions where hairpins have been inserted. The names of staple strands (as indicated by the gray circles) are determined based on the locations of these hairpins. For instance, the hairpin of stp_h4t2 can be found on helix no.4 and in the 2nd column of hairpins from the left. Binding sites for adapter strands and linker strands are highlighted with red dashed boxes. A red arrow indicates the point where the excess scaffold "loop" emerges from the origami structure. The DNA origami structure utilizes 2,688 bases from the p3024 scaffold. Out of the 336 unused base pairs, 316 are utilized to create a 6-helix bundle.

Building upon the foundational work of Mohammed *et al.*, we have refined the assembly of nunchuck structures by incorporating a shortened scaffold derived from a 3 kilobase ssDNA (p3024), as described by Nafisi *et al.*[45]. This modification significantly reduces the quantity of excess single-stranded DNA (ssDNA) that extends from the midpoint of each origami seed, a critical aspect of the design that ensures the structural integrity and functional fidelity of the nunchucks. The origami seeds are specifically designed to include regions that allow for the attachment of the linker strand at one end and adapter tiles at the other, facilitating the subsequent growth of the DNA nanotube arms. A detailed design scheme of the nunchuck seed can be found in Figure 2.2. Staple sequences and scaffold sequences can be found in the Supplementary Note of our previous publication [40].

Due to the supply scarcity of p3024 during the Covid pandemic, our laboratory had temporarily shifted to using a scaffold designated as p2717 for the assembly of nunchuck seed origami. This transition was supported by the development of an in-house pro-

tocol designed to replicate a precise 2,717-base pair segment from the p3024 plasmid. This process began with the amplification of the p3024 template through polymerase chain reaction (PCR), creating an abundance of double-stranded DNA corresponding to the desired 2,717-bp region. The subsequent step involved the deployment of alkaline denaturing gel electrophoresis, a technique chosen for its efficacy in dissociating the double-stranded DNA molecules into single strands. Through this method, we were able to selectively isolate the strand intended for use as the origami scaffold. The precision extraction from the gel was completed using the Zymoclean Gel RNA Recovery Kit. To further refine the scaffold's purity, a secondary purification phase was needed, utilizing a native agarose gel. This two-step purification process ensured the isolation of a high-fidelity single-stranded DNA, p2717, encompassing the full-length scaffold sequence needed for the folding of the nunchuck seed origami.

The protocol for p2717 production is summarized as follows:

1. Run one or several batches of PCR reactions to produce many copies of double-stranded p2717.

2. To maximize the amount of PCR product loaded onto the gel, use a SpeedVac vacuum concentrator to reduce the volume of the PCR products by as much as 70%. To denature the PCR product, add 2M NaOH solution to the product until the pH is around 12 as determined by a pH strip. After the PCR product sits at room temperature for about 20 min, add a dense gel loading dye containing extra glycerol to ensure the PCR product is heavy enough to sink into the wells. Our loading dye was made by mixing 24.4 mg bromophenol blue, 6.7 ml water and 27.7mL glycerol, and was used as a 6x solution.

3. Run the gel at 50V and 4°C (or on ice). For our 10-cm gel box, it took 17 h to bring the denatured bands to the bottom of the gel, which is recommended

for maximizing the separation of the two complementary strands. The exact time
needed for each gel box should be determined by experiment.

4. Wash the gel with 1x TAE buffer to remove urea, and stain with 0.01% ethidium
   bromide in 1x TAE for 1 h. An example of a stained gel illuminated by UV light
   is shown below in Figure 2.3 left. To determine whether the upper band or the
   lower band was the desired product, we extracted both and attempted to fold
   DNA origami with each, and concluded that the lower band was the desired p2717
   scaffold.

5. Carefully cut out the gel portion containing the lower band with a sharp and clean
   razor blade. Use the Zymoclean Gel RNA Recovery Kit (R1011) to extract the
   DNA from the gel piece.

6. The DNA extracted by the Zymo kit may contain some double-stranded DNA, since
   the two complementary strands in the denaturing gel were very close and cutting
   with a razor blade cannot precisely separate the lower band from the upper band.
   Any upper-band DNA remaining in the Zymo kit-extracted product will renature
   with lower-band DNA and become double-stranded p2717 again. Hence, we purify
   the DNA extracted from the Zymo kit once again by letting it sit in 1x TAE for
   10 min and running it on a native agarose gel. The native gel should contain 1.2%
   agarose in 1x TAE, and is pre-stained with 0.005% ethidium bromide. The gel is
   run at 120V for at least 30 min or until the p2717 scaffold band (see Figure 2.3
   right) is separated enough from the double-stranded p2717 band to be easily cut
   off with a razor blade.

7. Purify the p2717 scaffold with the Zymoclean Gel RNA Recovery Kit (R1011), use
   NanoDrop or other methods to determine its concentration, label and store it in
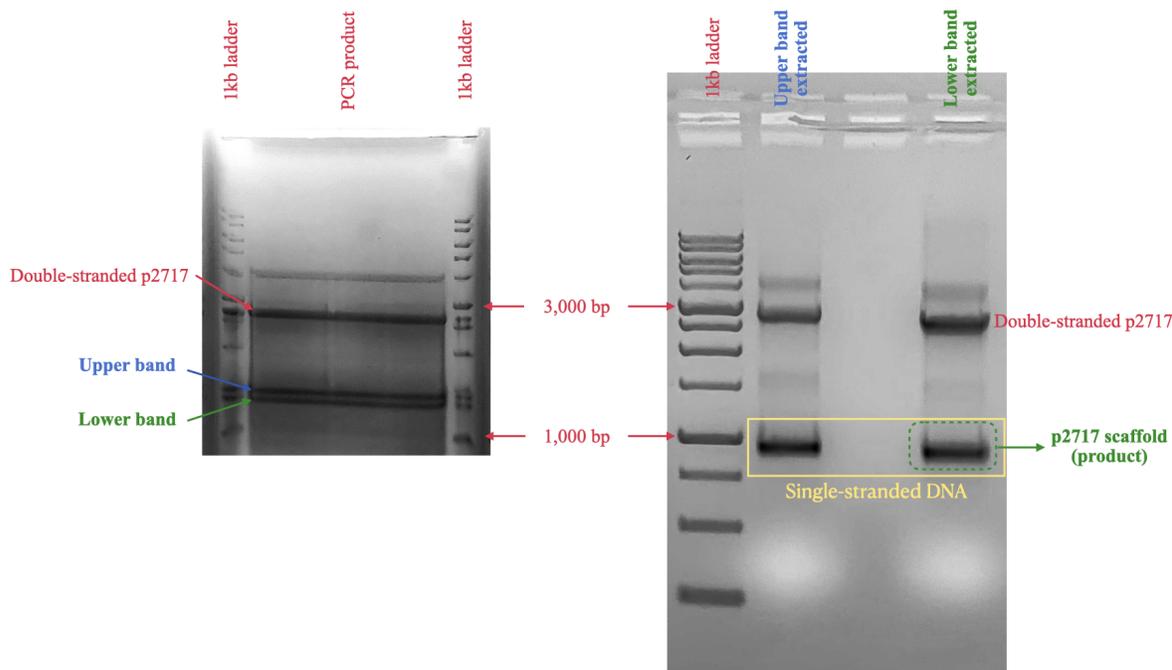
Figure 2.3:   Sample images for gels used in p2717 production. Left: the urea-agarose denaturing gel used to separate the two complementary strands of double-stranded p2717; right: the native agarose gel used to purify the single-stranded p2717 product by separating it from re-natured double-stranded p2717.

1x TAE buffer at $-20°$.

The methodology employed in the production of the p2717 scaffold is placed in the main body of this dissertation, rather than in the appendix, due to its potential significance beyond a mere procedural footnote in nunchuck assembly. The simplicity and accessibility of this technique, which utilizes readily available materials to produce kilobase length single-stranded DNA, could be valuable for researchers in the field of DNA origami. Custom ssDNA scaffolds are often a costly resource, while their double-stranded counterparts are more economically feasible. Therefore, the development of a reliable method to convert double-stranded DNA into ssDNA would be of considerable value to those exploring various ssDNA scaffolds for origami construction. However, though this method has successfully yielded the p2717 scaffold, our preliminary attempts to general-

ize the technique to other sequences have not been fruitful. Our initial efforts to denature other scaffolds, such as different 3-kb regions of M13mp18, did not result in the distinct band separation needed on urea-agarose denaturing gels, which may be attributed to the base composition of DNA[46]. At the moment, we document this method with the hope that it will inspire further research to refine and establish it as a standard for producing long ssDNA to be used for origami folding.

2.1.2   Arms and Adapters

Nunchuck arms consist of DNA DAE-E tiles, which are specially designed DNA structures that can nucleate and grow from the ends of the DNA origami seeds, extending the structure to form the "arms" of the nunchuck. The two arms of each nunchuck are made distinguishable by their incorporation of specific fluorophores, Cy3 and ATTO 488, which facilitates their identification and differentiation through fluorescence microscopy. In Figure 2.1(b), the REs and SEs tiles composing the nunchuck arms are depicted. Figure 2.1(c) offers a glimpse of the fully assembled nunchuck structure (IV). It is worth noting that these illustrations only partially represent the arms due to their extensive length. While the seeds are a mere 75 nm, the arms can reach up to a few microns, a length that far exceeds that of the seeds.

Adapters strands are designed with specific sequences that initiate the growth of REs and SEs nanotubes on either side of a nunchuck, essential for the intended asymmetry and functionality of the nunchuck's design. The design of adapter strands are shown in Figure 2.4, and sequences can be found in the Supplementary Note S3 of our previous work [40].
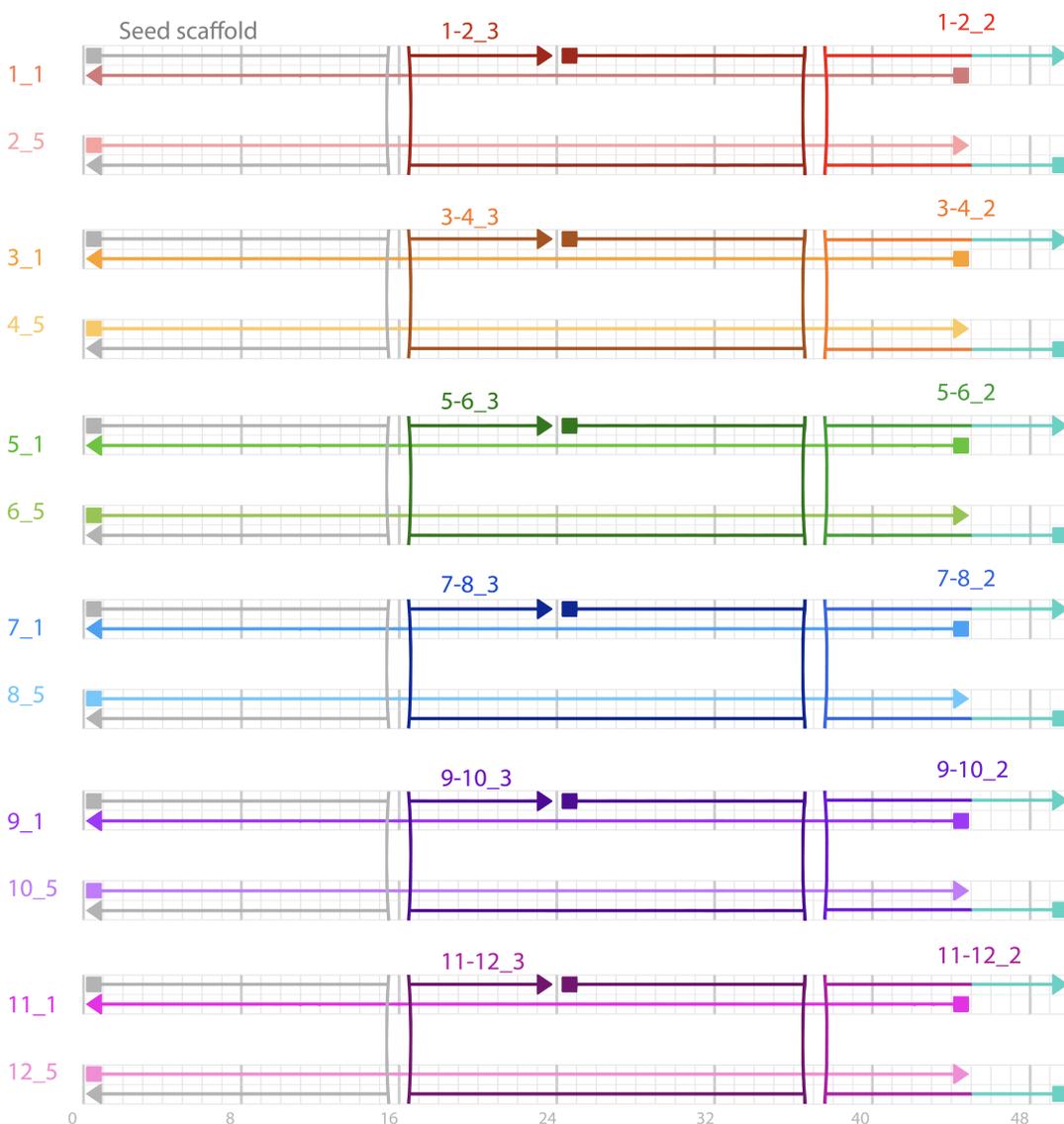
Figure 2.4:   This diagram details the arrangement of adapter strands on the seed scaffold, with the scaffold fringes depicted in grey. Each adapter strand is labeled according to its position on the scaffold helices (for example, "7_1" is located on helix number 7, while "11-12_3" spans helices 11 and 12) and their specific locations within each helical group. Adhering to the nomenclature established by Rothemund *et al.*[44], each group of adapter strands are shown in various shades of the same color and are sequenced as "_1", "_2", "_3", and "_5". "_4" is omitted because it is the scaffold strand. The overhangs of the "_2" strands are uniformly highlighted in turquoise to indicate their identical sequences across different helical regions.

2.1.3   Linker

The linker is the DNA strand that nunchucks are designed to measure and characterize. The design of these linkers is depicted in Figure 2.5. Linkers are essentially the duplex DNA segments that join two seeds, with their length being a defining factor for the relative positioning of the seeds around the linker axis due to the helical structure of DNA. We will later show that adopting a design strategy where the linker length corresponds to a half-integer number of helical turns (such as 37 bp instead of 32 bp) helps avoid steric hindrance and neutralizing any intrinsic bends at the seed-linker junctions. This consideration in design ensures the structural integrity and flexibility of the nunchuck. In most of our studies, as shown in Figure 2.5, these linkers contain nicks that are subsequently sealed through ligation to establish strand continuity. However, to evaluate the necessity of this step, we also explored a no-nick linker variant that eliminates the need for ligation; this alternative design will be briefly discussed in later chapters. Figure 2.6 shows cogli1 (https://sourceforge.net/projects/cogli1/) renderings of oxDNA simulations of the linker structures with and without the four T spacers at the linker-seed junctions[47]. Sequences for various linker designs, color-coded to distinguish different sections, are comprehensively listed in Supplementary Note S4 of our previous publication [40], providing a clear guide for replication and further study of these intricate nanostructures.

2.2   Nunchuck Assembly

2.2.1   Strand Preparation

The P3024 scaffold for our nunchucks was procured from Parsa Nafisi and Shawn Douglas at UCSF [45], with all other DNA strands synthesized by Integrated DNA Tech-
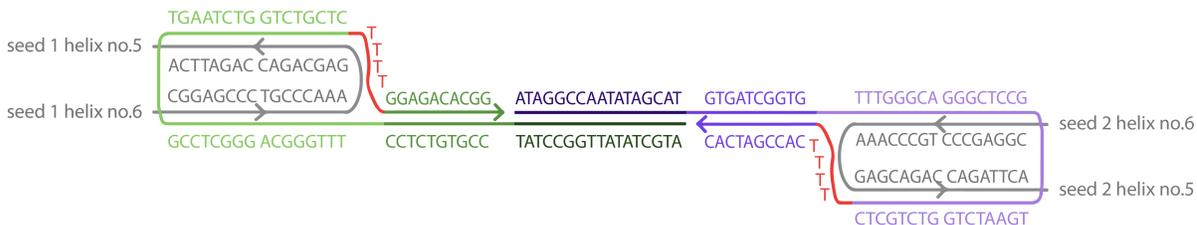
Figure 2.5: The design and sequences of a pair of 37 bp linker strands, illustrated by the green and purple segments, and their integration with the seed scaffold fringes, shown in grey. Within the linker strand, different shades denote distinct functional regions: the darkest shade indicates the pairing region to the complementary linker, the medium shade marks the self-hybridizing area, and the lightest shade shows the attachment site to the scaffold. The four red T bases at the linker-scaffold junctions act as spacers, and despite their presence, both oxDNA simulation and controlled measurements show that the T spacers cause no discernible differences in nunchuck flexibility (refer to Supplementary Note S15 of our previous publication [40]). Consequently, some of the linkers we studied did not have these T spacers.

Figure 2.6: Renderings of 58 bp linker structures with (top) and without (bottom) T-spacers simulated in oxDNA.

nologies. Strand purification was tailored to each type: staples underwent standard desalting, while linkers and adapters were HPLC or PAGE purified based on size. Linker strands were kinated to allow for subsequent ligation. For this, they were treated with T4 Polynucleotide kinase at 37°C for 2 h and then deactivated at 90°C. More details can be found in the "Methods" section of our previous publication [40].

2.2.2   Seed Formation

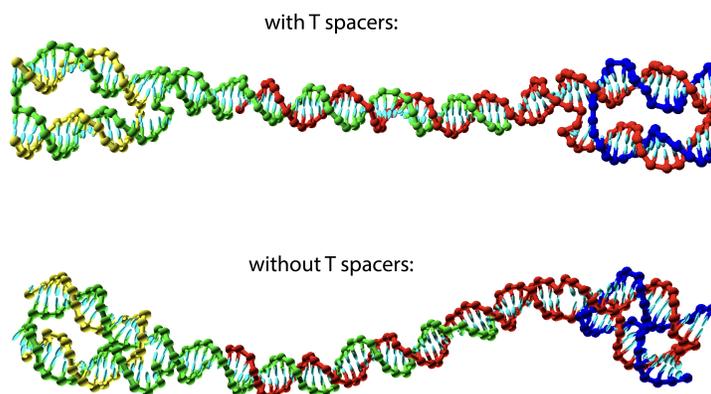To create a monomer seed, scaffold strands were mixed with various strands such as staple strands, adapter strands, and kinated linker strands in a TAE-Mg buffer. The mixture then followed a precise annealing profile involving a series of temperature decrements to enable proper folding and binding, as outlined in Supplementary Note S5 of our previous work [40]. Two monomers with complementary linkers were then mixed in equal volumes, and incubated at 31°C to form seed dimers. Post-annealing, the dimers undergo ligation using T4 DNA ligase and additional ATP at room temperature, followed by an incubation with proteinase K at 37°C to deactivate the ligase. This last step is crucial for creating stable, dimerized seeds that are integral to the nunchuck framework.

2.2.3   Seed Purification

To ensure the purity of our nunchuck dimers, we employ agarose gel electrophoresis. The process begins by casting 1% high-resolution agarose gels, into which the ligated seed dimer mixture is loaded. During electrophoresis, the dimers, which migrate more slowly than monomers due to their size, are separated out. Post-electrophoresis, distinct bands corresponding to the dimers are excised from the gel. These fragments are then extracted using the Freeze N Squeeze DNA Gel Extraction Spin Columns (Bio-Rad, 7326166) and stored at room temperature, ready for use in subsequent nunchuck assembly steps within
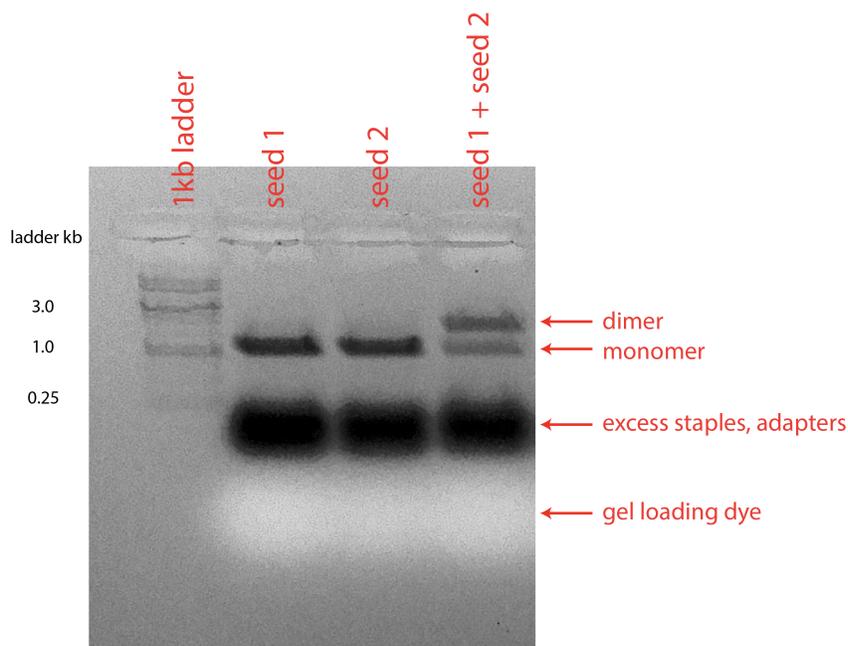
Figure 2.7:   A sample agarose gel showing EtBr-stained nunchuck prepara-
tions. The gel illustrates distinct bands corresponding to the monomers and dimers
of nunchuck seeds, with dimers appearing in Lane 4 with less mobility due to their
larger size. The fastest-migrating and brightest bands, present in Lanes 2, 3, and
4, represent excess staples and adapter strands that are separated from the desired
constructs. In Lane 4, the clear separation of band sizes across the lanes demonstrates
the effective purification process.

two weeks. A sample gel image for seed purification is shown in Figure 2.7.

2.2.4   Nunchuck Formation and Yield

The assembly of nunchucks involves the careful annealing of REs and SEs tile strands,
which are mixed in equimolar ratios. The annealing protocol begins at 90°C and slowly
decreases to 31°C, pausing at 40°C for 10 min to incorporate the seed dimers, where the
purified dimers are manually added to the annealing solutions in the thermocycler. To
mitigate the risk of DNA adherence to the PCR tube walls during the formation process,
BSA solution is added. The specifics of this procedure are detailed in the "Methods" sec-
tion of our previous publication [40], with an example recipe presented in Supplementary
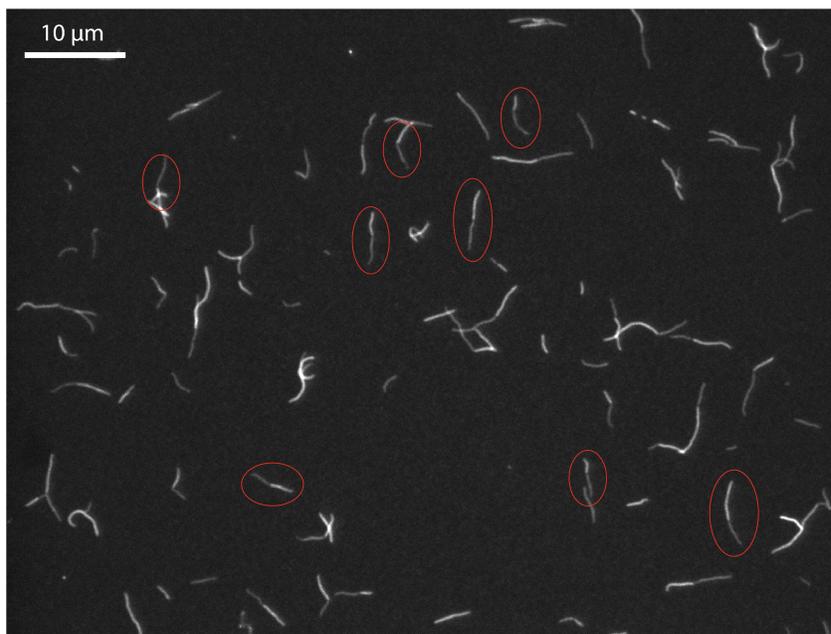
Figure 2.8:   This image presents a high-yield nunchuck sample. The formation of these structures and the imaging procedures are detailed in the "Methods" section of our previous work[40], specifically under "Nunchuck Formation". Here, "yield" is defined as the number of nunchucks divided by the number of all tube-like structures (including nunchucks and spontaneously-formed DNA nanotubes). The distinguishing feature of nunchucks, the "bright arm-dim arm" configuration, which is visible in this image, facilitates their identification under a microscope. Selected nunchucks have been highlighted with red circles to exemplify this characteristic. In our most successful samples, the yield often surpasses the 50% mark.

Note S5. After the annealing process, the nunchucks are incubated overnight at 31°C, which allows them to grow to a desired length.

Yield, a key metric of success in nunchuck synthesis, is quantified by the ratio of actual nunchucks to all tubular structures observed in the annealed sample, including both target nunchucks and any spontaneously formed DNA nanotubes. Figure 2.8 shows a high-yield sample with numerous nunchucks, identifiable by their unique fluorescence features, with some highlighted by red circles. In optimal conditions, yields often surpass the 50% threshold, indicating a highly efficient assembly process.

2.3   Nunchuck Imaging

2.3.1   Preliminary Imaging

While selecting for nunchucks that are most suitable for imaging, we faced the challenge of unpredictable arm lengths. Nunchuck arm length is a crucial parameter: too short, and it hampers our ability to determine bend angles from images; too long, and it not only becomes difficult to image but also slows the diffusion of the nunchucks, which is undesirable for our dynamic studies, where time resolution is key. To address this, when preparing nunchucks for imaging, we anneal several nunchuck samples with different amounts of seeds, as the ratio of seed to tile directly affects the arm length and the overall yield of nunchucks compared to spontaneously nucleated nanotubes. We then conduct a series of preliminary imaging trials, mixing a small volume of each annealed nunchuck solution with TAE buffer on a coverslip, sealing it to a glass slide with candle wax, and examining the sample under a fluorescence microscope. Figure 2.8 shows an example image taken at this step. These trials were to empirically determine which sample provided nunchucks with the most suitable arm length for detailed, dynamic imaging, where the linker mechanics could be studied. Since no image data from this step is to be analyzed, it is not necessary to clean the coverslips and glass slides ahead of time.

2.3.2   Dynamic Imaging

In order to visualize the end-to-end bend angle of linkers as the nunchucks freely diffuse in a solution, our imaging strategy involves subjecting nunchucks to quasi two-dimensional confinement, which is provided by a narrow gap between a coverslip and a glass slide. When moving in a plane, their bend angles can be observed and recorded accurately, as the perpendicular fluorescence imaging eliminates the complexity of three-

dimensional tracking and the potential for visual distortion. The cleanliness of these surfaces is critical, as any impurities could affect the nunchucks' constrained movement and the clarity of fluorescence images used for angle measurement.

To ensure an optimal imaging environment, glass slides and coverslips undergo a meticulous cleaning protocol. They are first sonicated in a solution of ultrapure water with a small volume of concentrated cleaning solution to remove organic and inorganic residues. After sonication, they are thoroughly rinsed with pure water several more times to eliminate any residue from the cleaning agents. The glassware is then soaked in 2M NaOH to etch away a thin layer, ensuring a clean surface. After another round of rinsing to remove the NaOH, a final wash with ethanol is performed to displace water, enabling faster drying. The glass is dried in a 50°C oven and stored in clean jars to maintain its pristine condition until use. This process ensures that the glass surfaces are free from impurities that could affect imaging quality.

To keep nunchucks from potential interactions with the glass slide and coverslip, which affects their bending, we passivate the glass surfaces after cleaning and before using to construct samples. To achieve inert glass surfaces, a solution of 300 μL of ethanol, 3 μL of glacial acetic acid, and 2.5 mg of mPEG-silane is prepared. mPEG-silane, insoluble in ethanol alone, requires heating for 5 min at 50°C to dissolve properly, forming the passivation mixture. 15 μL of this mixture is then deposited onto a clean coverslip, which is immediately inverted onto a clean slide. This action allows the solution to evenly spread and fill the gap between the slide and the coverslip, creating a uniform passivation layer. The slide-coverslip pair is then placed in a sealed environment saturated with ethanol vapor to prevent the mPEG-silane solution from drying out prematurely. After a curing period of three days, the glassware is ready for use and retains its passivation properties for at least two months. Through this process, we ensure that the glass surfaces are conducive to observing the intrinsic bending behavior of nunchucks without

the complication of surface-induced artifacts.

To ensure that the nunchuck remains within the plane of focus during imaging and that the observed bend angle accurately represents the linker's end-to-end bend angle, the gap between the passivated coverslip and slide surfaces must be minimized. This narrow confinement is crucial to prevent any out-of-plane movements that could distort the imaging results. For this purpose, we use the custom-built jig, depicted in Figure 2.9, which is used to precisely align and press the PEG-coated glass surfaces together. The design of the jig, with its robust aluminum base, adjustable screws, and wing nuts, allows for controlled pressure to be applied uniformly across the glass surfaces. The incorporation of Teflon cylinders and PTFE/silicone septum ensures that the pressure is distributed evenly without damaging the thin coverslip. While the glass pieces are positioned and secured within the jig, UV-curing nail polish is applied along the edges to bond the coverslip to the slide. After curing, the flow cell is carefully removed from the jig, resulting in a sealed environment with openings at two opposite ends. These openings are crucial for introducing the sample into the flow cell.

Before loading nunchuck solution into the flow cell, a PCA-trolox oxygen scavenging system (OSS) was introduced to the nunchuck sample, to mitigate the photobleaching of fluorescent tags during microscopy[48]. For each 1 μL of the nunchuck sample, we add 1 μL PCD solution and 1 μL PCA-trolox solution, where the PCD solution consisted of 1 μL of PCD (originally at 1 μM concentration) mixed with 2 μL of 1x TAE buffer (containing no magnesium), and the PCA-trolox solution consisted of 0.75 μL of PCA at 67 mM, mixed with 0.25 μL of trolox at 83 mM and 4 μL of 1x TAE buffer (containing no magnesium). The mixture was then loaded onto the flow cell immediately before imaging, as the reaction starts once the PCD solution contacts the PCA-trolox solution. For imaging, a 0.6 μL droplet of this nunchuck-OSS mixture was placed onto one opening of the freshly assembled flow cell. Capillary action naturally draws the solution across the
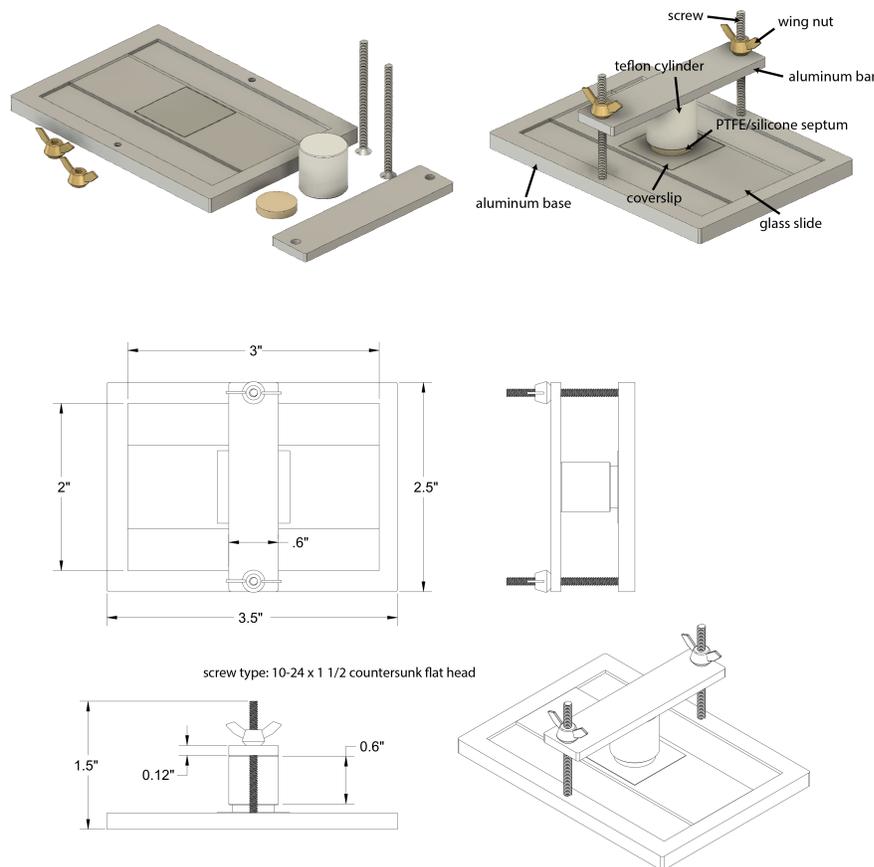
Figure 2.9:   The depicted jig is designed for pressing PEG-coated coverslips against the glass slide. The mechanical drawings detail the assembly and dimensions of the jig, which is operated by manually tightening the wing nuts until they reach a natural stop, ensuring a uniform application of pressure.

narrow gap between the coverslip and slide, ensuring even distribution. Once the solution reaches the far end of the flow cell, excess fluid is gently wiped off, and both ends are promptly sealed with candle wax to prevent leakage and evaporation. It's noted that the process of populating the flow cell tends to decrease the observable concentration of nunchucks, with a typical sample displaying approximately one nunchuck per ten nanotubes.

Imaging was conducted using an Olympus IX70 inverted microscope in epi-fluorescence mode, equipped with a 100x/1.30 oil immersion objective and a Cooke Sensicam QE cam-

era. The fluorescent molecules were excited and imaged using specific optical filters: for Cy3, a 325AF45 excitation filter, a 560DRLP dichroic mirror, and an HQ610/75 emission filter; and for ATTO 488, a 475DF40 excitation filter, a 505DRLP dichroic mirror, and an HQ535/50 emission filter. Criteria for selecting nunchucks for imaging included: arm lengths between 2 and 4 microns without visible defects, sparse surrounding tubes to avoid crowding, the ability of nunchucks to move and bend freely within the observed area, and a focus that remained sharp across all observations. Image sequences were typically captured with intervals bewteen 0.1 (for continuous imaging) and 3 seconds (for decorrelated imaging) between frames, with an initial exposure time of 10 milliseconds per frame. This exposure was gradually increased over time for long sequences where photobleaching became apparent after more than 1,000 frames. We have established, as detailed in the next section, that the bend angle decorrelation time is approximately 1 second. Consequently, the 3-second interval between frames was adopted to ensure definite decorrelation between neighboring frames, although the 1-second intervals were used more often. Depending on characteristics of the samples such as the freshness of the OSS solutions and the level of airtightness of the seal, we were often able to monitor a nunchuck for thousands of frames. The total number of frames of images we were able to collect for each nunchuck was usually unrelated to the frame rate. This is because the samples were illuminated only during the camera's image acquisition phase, with the shutter effectively blocking the light at all other times, thereby preventing continuous exposure. Imaging sessions concluded either upon achieving the desired number of frames or if the nunchucks began to show signs of degradation, such as breaks or kinks in the arms, or if they adhered to adjacent structures or the glass surface.

To prepare the images for analysis, they underwent a contrast enhancement process to sharpen the distinction between the nunchucks and the background. Subsequently, each image was cropped to a 200 ×200 pixel square, precisely centered around the bounding

box containing the nunchuck object, ensuring that the focus was solely on the nunchuck of interest for bend angle measurement.

2.4   Nunchuck Angular Decorrelation Time

2.4.1   The Nunchuck Angular Diffusion Coefficient

To identify an optimal sampling frequency that would yield independent bend angles in our observations, we conducted a detailed analysis of the fluctuation patterns in the bend angles of nunchucks, specifically those with the longest linkers in our study, measuring 58 bp. These were imaged at a high temporal resolution of 7.6 Hz, from which we derived the time series $\theta(t)$. The challenge in determining a precise decorrelation time stemmed from the indistinct changes in step size, $\Delta(\delta t) \equiv \theta(t + \delta t) - \theta(t)$, especially pronounced near the angles $\theta = \{0°, 180°\}$, a topic we delve into later in this section. To circumvent this issue, we computed the angular diffusion coefficient, $D_\theta \equiv \Delta^2/(2 \cdot \delta t)$, using successive frames. This calculation specifically employed frames where the bend angle was increasing, beginning from 20° to 90°, as well as those where it was decreasing, beginning from 50° to 120° (refer to Figure 2.10(a), illustrated by blue and green lines, respectively). These intervals were strategically chosen to base our calculations of $\Delta(\delta t = 1/7.6\ s)$ on angles that were comfortably distant from the ambiguous $\theta = \{0°, 180°\}$ and that were most representative of the angles predominantly sampled due to thermal fluctuations, lying within 2σ of the equilibrium distribution. Distributions of Δ calculated from qualified consecutive frames fit well to a folded normal distribution with $\mu = 0°$ (Figure 2.10(b)). D_θ decreased with increasing nunchuck arm length, as expected (Figure 2.10(c)).
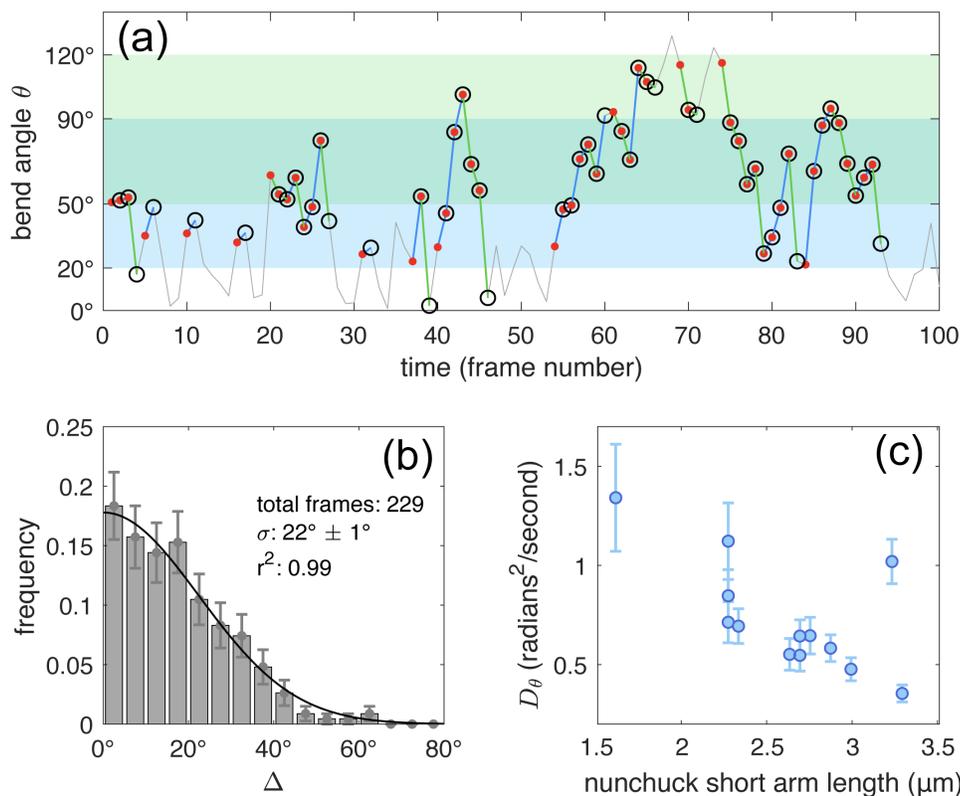
Figure 2.10: (a) This panel presents a time series analysis of the bend angle (θ) for a nunchuck with a 58 bp linker, captured at a frequency of 7.6 Hz. The differential bend angle $\Delta = \theta_{i+1} - \theta_i$ is calculated from positive steps (indicated with blue lines) beginning within a bend angle between $20°$ and $90°$ (shaded in blue), and negative steps (indicated by green lines) originating between $50°$ and $120°$ (shaded in green). These ranges were selected to minimize the ambiguity at critical angles ($\theta = \{0°, 180°\}$) and to mitigate any effects from non-thermal forces. The initiation and termination points of steps meeting the criteria are denoted by red dots and black circles, respectively, while steps not meeting the criteria are represented with grey lines. (b) A sample distribution of Δ, which has been modeled with a folded normal distribution (black curve). (c) The angular diffusion coefficient (D_θ), indicative of rotational movement, for 13 distinct nunchucks, each with a 58 bp linker, is plotted against the length of their shorter arm. This coefficient characterizes the rate at which the bend angle of the nunchucks changes over time, providing insights into their dynamic behavior.

2.4.2   Interpreting the Angular Diffusion Coefficient

The calculated angular diffusion coefficients, D_θ, imply that to ensure the bend angle undergoes a complete 180° transition between frames — a robust benchmark for ensuring decorrelation — one would need to employ a sampling frequency within the range of 0.07 Hz to 0.25 Hz. Regrettably, such low-frequency sampling proved to be incompatible with our manual tracking system for individual nunchucks. Conversely, imaging at a frequency of 1 Hz allowed for the consistent tracking of single nunchucks up to the point of photobleaching. Furthermore, when imaging or downsampling at a frequency of $0.\bar{3}$ Hz, we observed histograms that were statistically indistinguishable from those obtained at 1 Hz, suggesting that our ample sample sizes sufficiently counterbalanced most of the effects stemming from any lingering correlation. The measured values, 0.35 rad^2/s $\leq$ $D_\theta \leq 1.35$ rad^2/s, align closely with the theoretical diffusion coefficient predicted for a prolate ellipsoid characterized by a semi-major axis (a) of 2 μm and a semi-minor axis (v) of 5 nm, rotating in water at a viscosity of $\eta = 8.9 \cdot 10^{-10}$ pN·s/nm^2 and at room temperature, estimated at $k_B T = 4.1$ pN·nm [40]:

$$D_\theta = \frac{k_B T}{f_r} = \frac{3 k_B T \left(\ln \frac{2a}{b} - \frac{1}{2} \right)}{8 \pi \eta a^3} = 0.43 \, \text{rad}^2/\text{s} \tag{2.1}$$

The decline in sensitivity to arm length, as observed in our data, was less pronounced than the theoretical model would predict; we noted a linear diminution rather than the expected cubic relationship with respect to arm length (Figure 2.10(c)). Currently, the underlying cause for this observed pattern remains unclear. While other researchers have delved into the solution dynamics of macromolecules with flexible segments[49], focusing particularly on the angular diffusion of hinge-like structures, the specific relationship between a hinge angle's diffusion coefficient and the lengths of its arms, even from a two-dimensional perspective, seems to be an area that has not yet been thoroughly explored.
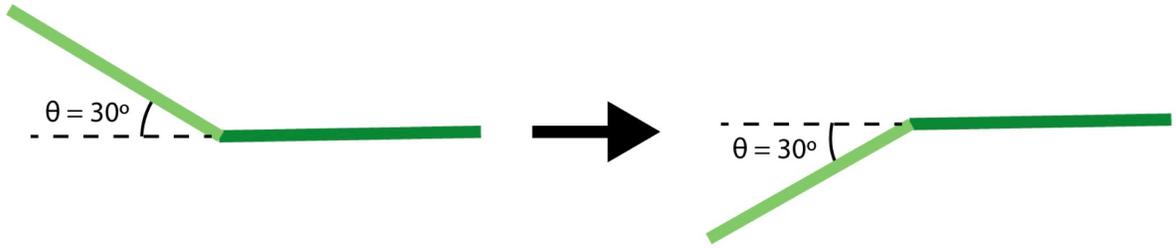
Figure 2.11:   The cartoon drawing shows a nunchuck undergoing a 60° alteration in bend angle, demonstrating a scenario where the change is not reflected by the value of $\Delta(\delta t) \equiv \theta(t + \delta t) - \theta(t)$, which remains at 0°. This highlights the limitations of $\Delta(\delta t)$ in accurately capturing the dynamics of nunchuck movement near extreme angles.

2.4.3   Ambiguity in Bend Angle Changes

Figure 2.11 illustrates a nunchuck experiencing a 60° shift in its bend angle; however, the calculated $\Delta(\delta t) \equiv \theta(t + \delta t) - \theta(t)$ for this movement registers as 0°. This discrepancy indicates that Δ_θ fails to fully represent the actual physical alteration in the nunchuck's bend states when θ approaches the extreme values of 0° or 180°. To circumvent this issue, our approach for computing angular diffusion coefficients was to selectively use steps that started and ended at angles distinctly away from $\theta = \{0°, 180°\}$, thereby reducing potential misrepresentations.

2.4.4   The Nunchuck Angular Decorrelation Time

Our next objective is to define the decorrelation time for nunchucks. This will enable us to establish the requisite temporal interval between successive frames to accumulate a dataset that is sufficiently independent for the construction of a representative histogram. To this end, we must characterize the autocorrelation function of the bend angle, denoted

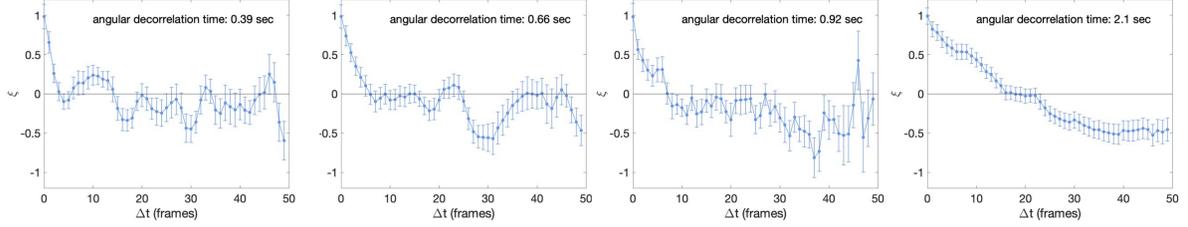Figure 2.12:   plots of $\xi(\Delta t)$ vs. Δt (in the unit of frames) of 4 example image sequences. The angular decorrelation time, defines as the time when the correlation first drops to 0 within uncertainty, is shown in each subplot.

as $\xi(\Delta t)$, for a given image sequence, defined by the following expression:

$$\xi(\Delta t) = \left\langle \frac{(\theta(t) - \overline{\theta}) \cdot (\theta(t + \Delta t) - \overline{\theta})}{\sigma_\theta^2} \right\rangle \tag{2.2}$$

Here, Δt represents the interval between any two consecutive frames, measured either in seconds or as the number of frames. The variable $\overline{\theta}$ is the average bend angle, computed over all values of $\theta(t)$ or $\theta(t + \Delta t)$ included in the summation. The term σ_θ denotes the standard deviation of all the bend angle measurements that are part of the summation. This autocorrelation function, $\xi(\Delta t)$, can be derived from image sequences captured at a rate of 7.6 frames per second, under the condition that the nunchuck maintains a bend angle within the range of 10° to 170°. This specific range ensures that potential ambiguities at $\theta = 0°$ and 180° do not skew the results. We computed the values of $\xi(\Delta t)$ for 42 such image sequences, considering Δ_t values extending to 50 frames. Plots of $\xi(\Delta t)$ for these calculations are presented in Figure 2.12.

The angular decorrelation time is determined by the point at which the correlation, $\xi(\Delta t)$, diminishes to zero, allowing for the margin of uncertainty. During our analysis, we noted that certain image sequences exhibited intervals where the nunchuck traversed various ranges of bend angles, indicative of different bend states. In these instances, the resulting decorrelation time, derived from the global average, was significantly over-
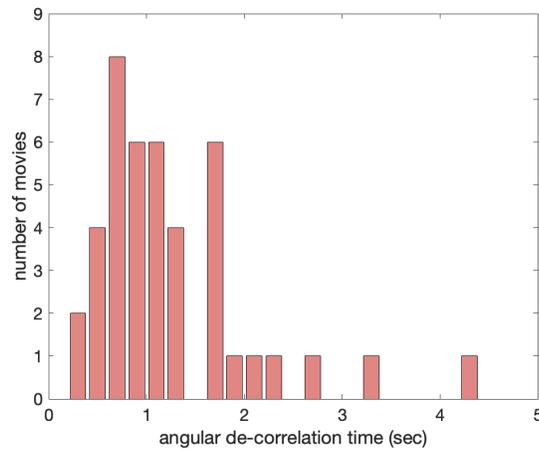
Figure 2.13: The distribution of angular decorrelation time calculated from 42 image sequences during which nunchucks remained away from boundary conditions.

estimated. Consequently, while the decorrelation time exhibited considerable variation across different nunchucks (as summarized by Figure 2.13), it was typically on the order of 1 second.

In light of these findings, we concluded that a 1-second interval between consecutive frames would be sufficient to generate bend angle histograms without significant corre-lation effects. To err on the side of caution, we have sometimes opted for a 3-second interval, although this extended duration markedly increased the time required for imag-ing. Consequently, a 1-second interval was more frequently employed.

Chapter 3

Nunchuck Image Analysis

In the initial stages of our research, we employed filament tracing to map the contours of nunchucks and determine their bend angles. This method, as shown in part (e) of Figure 3.1, involved delineating the nunchuck's outline using the JFilament plugin for ImageJ. To illustrate this process, the contour in the example is traced in yellow, with the location of the nunchuck vertex highlighted by a red dot. From this point, measurements were taken along the arms, quantifying the degree of bending by calculating vector dot products. Although this process was highly accurate, it was also labor-intensive, requiring continuous manual input and extensive mouse clicking to adjust the filament, which proved to be physically taxing over time. This repetitive manual interaction led to ergonomic issues, such as shoulder and arm strain, a problem I personally experienced, along with my colleagues, after prolonged periods of analysis. Consequently, we made the decision to switch to a neural network approach for analyzing bend angles. The transition was motivated not only by health considerations but also by the need for a more efficient method to handle the high volume of data. The neural network allowed us to automate the analysis process, significantly reducing the physical strain associated with continuous mouse use. Furthermore, it proved to be a much faster method, capable

of processing complex images with greater consistency. It adeptly handled typical imaging non-idealities, such as fluctuations in contrast or the interference of adjacent nanotubes, streamlining our workflow and contributing to the overall advancement of our research.

3.1   The Neural Network

In this section, we introduce our in-house convolutional neural network (CNN) named NunchuckNet, tailored specifically for the analysis of nunchuck bend angles from their fluorescence images. CNNs are a class of deep learning algorithms that have revolutionized image recognition and analysis. Their structure is inspired by the organization of the animal visual cortex and is particularly adept at capturing spatial hierarchies in imagery. CNNs consist of multiple layers of processing, including convolutional layers that automatically and adaptively learn spatial hierarchies of features from images. These features are then used by fully connected layers to perform the classification task. NunchuckNet represents a significant leap forward from manual filament tracing methods, offering a rapid and accurate classification of nunchuck images into discrete 5° bend angle bins, ranging from $-180°$ to $180°$. The training of NunchuckNet was an extensive endeavor, utilizing 570,960 computer-generated images that captured a variety of scenarios akin to experimental conditions. These images included common complexities such as curved arms and variable contrast, as well as potential obstructions like adjacent nanotubes. By encompassing these "non-idealities", the network learned to navigate the nuanced landscape of real nunchuck images, discerning bend angles with a high degree of accuracy. The structure of NunchuckNet is based on the AlexNet architecture, a pioneering CNN framework consisting of eight layers[50]. The initial five are convolutional, designed to extract a hierarchy of visual features, and the final three are fully connected layers, culminating in a classification output. In our implementation, the output layer consists
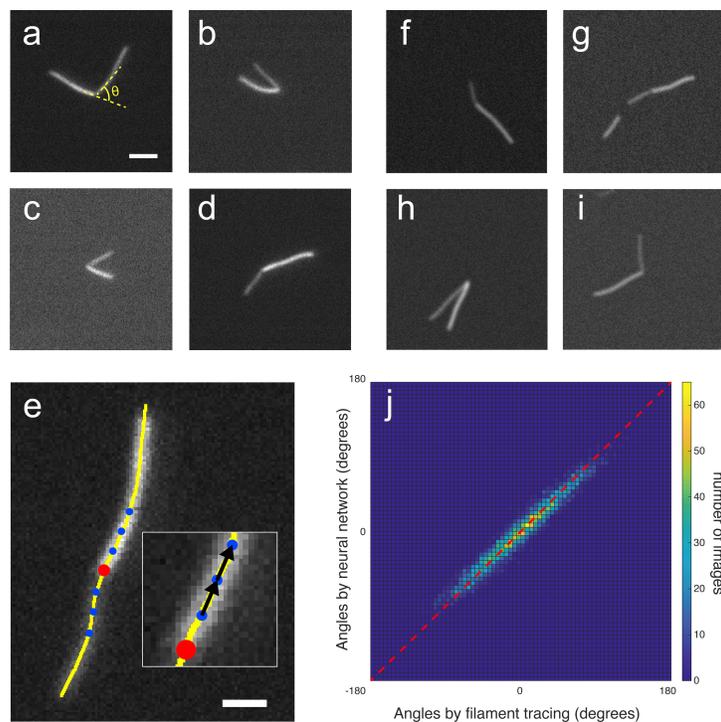
Figure 3.1: Selected nunchuck images and visualization of nunchuck bend angle determination techniques. (a-d) display a series of nunchuck images captured under fluorescence microscopy, scaled uniformly (scale bar = 2 μm). A discernible difference in brightness between the two arms of each nunchuck can be seen. This brightness difference is due to the differing amounts of fluorophores each arm carries, which allows for clear differentiation between the two. This contrast facilitates the precise identification of the linker's location, a vital aspect of analyzing the nunchuck's structure and bend angles. The defined bend angle, θ, ranges from 0° for straight nunchucks to 180° for fully bent configurations. (e) Bend angle measurements were initially conducted via the filament tracing method, where the nunchuck contour is traced in yellow using the JFilament plugin in ImageJ. The nunchuck vertex, marked by a red dot, is determined by a change in brightness along the nunchuck's contour. Beginning 8 pixels from this vertex, vectors are plotted along the arms, which results in 64 dot products from each pair of 8 vectors on either side of the vertex (for instance as demonstrated by the vector arrows in the inset), aiding in the determination of θ and its variance (scale bar = 2 μm). (f-i) depict computer-generated nunchuck images, which were employed to train the neural network. These simulations incorporate typical image aberrations, like arm curvature, contrast variability, and proximity of other nanotubes, to closely resemble actual imaging conditions. (j) illustrates a confusion matrix that compares the bend angles derived from 5,314 real nunchuck images via both filament tracing (horizontal axis) and neural network analysis (vertical axis), highlighting the comparative accuracy of the two methods.

of 72 classes, each corresponding to the center of a 5° bin. This granularity allows for a refined analysis of the nunchuck's bend angle, with each class representing a distinct orientation such as 2.5° or −117.5°.

The utility of NunchuckNet is twofold; it not only enhances the efficiency of image analysis by classifying angles rapidly but also improves precision by mitigating the subjective inaccuracies inherent in manual tracing. The neural network's training and subsequent application were conducted using MATLAB's Deep Learning Toolbox, a friendly environment for developing sophisticated deep learning models. The codebase for NunchuckNet, along with the neural-network related MATLAB programs, is openly available at https://github.com/caixinyue/nunchuck.

3.1.1   Training Set Generation

In the field of convolutional neural networks, the use of synthetic training data has emerged as a viable solution to overcome the challenges often encountered with real datasets. This approach has seen success in various domains, such as object detection, and offers several advantages for training neural networks like NunchuckNet. By utilizing computer-generated images, we circumvent the limitations of small, hand-labeled datasets, eliminate the potential bias from unevenly distributed training samples, and ensure absolute certainty in image classification - advantages that are particularly beneficial when real data is scarce or laborious to label. Importantly, most nunchucks seldom form a bend close to 180°, resulting in a scarcity of such images, however, for effective training, a substantial number of these rare images is essential. This issue can only be overcome by using synthetic training data.

For NunchuckNet, training images were algorithmically created in MATLAB, simulating the properties of actual nunchuck images. The process began with the random
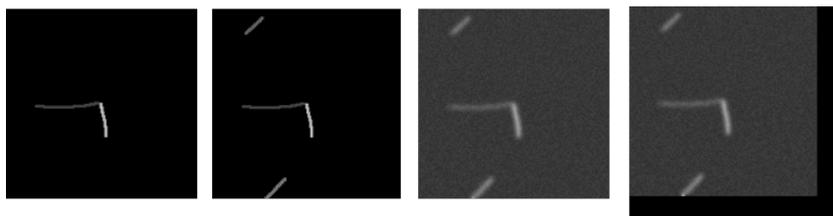
Figure 3.2:   Synthetic training images for NunchuckNet. The images display comput-er-generated nunchucks with randomized vertex placement and arm lengths within a 200 px × 200 px field, resembling the variability in real nunchuck images. The left-most image shows an isolated nunchuck, while subsequent images introduce additional complexities, such as background nanotubes and noise. The rightmost image is used to train the NunchuckNet.

placement of a nunchuck's vertex within a predefined central region of a blank canvas. Arms, represented by stochastic paths, were then generated from this vertex point, with lengths and initial brightness levels randomized within realistic parameters. To mimic the fluorescent disparity seen in experimental images, the second arm was assigned a brightness level approximately half of the first, introducing a controlled variability into the training data. The arms themselves were shaped by a random walk algorithm, fine-tuned to reflect the 2D persistence length characteristic of DNA nanotubes, and subse-quently smoothed to eliminate any unnatural sharp bends. This generation process also included the creation of background DNA nanotubes, although images with these ele-ments interfering with the nunchuck's vertex were excluded, since such nunchuck images were never analyzed in reality. To further enhance the realism of these synthetic images, a blur effect, simulating a 2D heat diffusion process, was applied, along with the addition of random noise, thereby approximating the typical visual artifacts of actual microscopy data. The final stage of image preparation involved resizing to fit the input requirements of the AlexNet architecture used in NunchuckNet. The images were expanded with black borders to the required dimensions (227 px × 227 px) and converted from grayscale to RGB format.
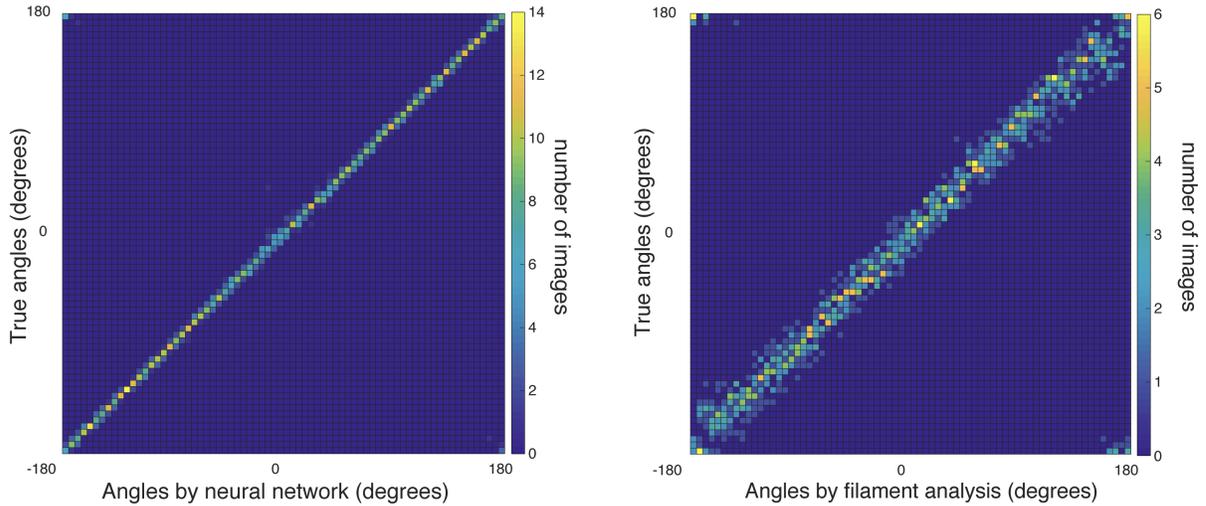
Figure 3.3: Confusion matrices showing the comparison of true bend angles with those determined by NunchuckNet (left) or by filament analysis (right) for a set of 1,008 computer-generated nunchuck images. The vertical axes represent the true angles, while the horizontal axes display the angles as measured by NunchuckNet or filament analysis, with the color intensity indicating the number of images per bin. The concentration of data along the diagonal indicates a high level of precision in angle determination by the neural network.

3.1.2 Neural Network Training and Performance Evaluation

NunchuckNet was exposed to a vast array of 570,960 computer-generated images, ensuring an even distribution across the 72 angle classes. The training process employed the stochastic gradient descent with momentum (sgdm) optimizer, with a starting learning rate of 0.01, which was reduced tenfold after each epoch. To avoid overfitting and ensure generalization, the images were shuffled after each epoch, and the network was validated against a separate set of 144,000 images. The training was monitored and concluded after 4655 iterations, approximately 3.5 epochs, when the validation accuracy stabilized at 53.3%.

To assess NunchuckNet's accuracy and to benchmark it against traditional filament analysis, we generated a new set of 1,008 computer-generated images with evenly distributed bend angles. Both the NunchuckNet and the traditional filament tracing method

37

were used to analyze these images. The performance comparison, visualized in Figure 3.3 through confusion matrices, demonstrates NunchuckNet's superior precision. The concentrated diagonal in the neural network's confusion matrix signifies a high level of accuracy, with 63.3% of images correctly classified and only 2% deviating by more than one bin from the true angle — a testament to the network's efficacy despite the seemingly modest training accuracy. Moreover, NunchuckNet offers a significant advantage in terms of efficiency and user ergonomics. While filament tracing is laborious, requiring intensive manual input and approximately an hour to analyze a 500-frame sequence, NunchuckNet can process the same sequence in about a minute without any user intervention. This dramatic reduction in analysis time and the elimination of repetitive tasks safeguard researchers from strain injuries associated with manual analysis methods.

3.1.3   Application on Nunchuck Images

When applied to real image data, NunchuckNet operates by evaluating the likelihood of each image belonging to one of the 72 bend angle classes. It assigns a confidence score to each class and identifies the bend angle based on the class with the highest score. The robustness of this approach is evident in Figure 3.4, where several real nunchuck images are analyzed, with NunchuckNet providing both the bend angle and its corresponding likelihood score (being a probablity, the maximum value for thi score is 1). Following the NunchuckNet analysis, we implement a stringent quality control measure where images with scores falling more than two standard deviations below the sequence mean are considered unreliable and are thus rejected. The bend angles for these frames are categorized as unknown and are not included in the final reported results, such as histograms or statistical analyses. This step is crucial in maintaining the integrity of our data, ensuring that only frames with a high confidence prediction contribute to our findings.

A noteworthy observation from NunchuckNet's performance on real data is its conservative scoring for nunchucks with extremely large bends, specifically those with absolute angles greater than 170°. Such configurations increasingly resemble straight nanotubes, leading to lower confidence scores, as demonstrated in image (h) of Figure 3.4. Simply discarding these images could significantly skew the accuracy of bend angle statistics, particularly in cases where nunchucks have inherently bent linkers prone to high-bend conformations. To mitigate this, we have implemented a manual review protocol. Each image flagged with a large bend angle by the neural network, regardless of its confidence score, undergoes a manual validation process. This is essential to prevent the inadvertent exclusion of valid, highly-bent nunchuck images and to maintain the integrity of our statistical analysis. To streamline this process, we have developed a customized MATLAB program specifically for this task. This tool expedites the review by presenting each suspect frame to the human evaluator, allowing for a swift and accurate determination of the appropriateness of the neural network's bend angle classification. Images confirmed as accurate are retained for inclusion in the dataset, whereas those deemed incorrect are removed.

3.1.4   Improving the Neural Network

In our ongoing efforts to refine the analysis of nunchuck bend angles, we have built upon the foundations laid by the NunchuckNet (hereater referred to as NN1 for clarity), which was published in our previous work [40]. Since then, we have progressed beyond our initial NN1, and developed more advanced iterations of the neural network — NN2, NN3, and NN4 — each surpassing its predecessor in terms of training image quality and feature representation. As described in previous sections, while NN1 significantly reduced analysis time and outperformed traditional methods in accuracy, there remained flaws
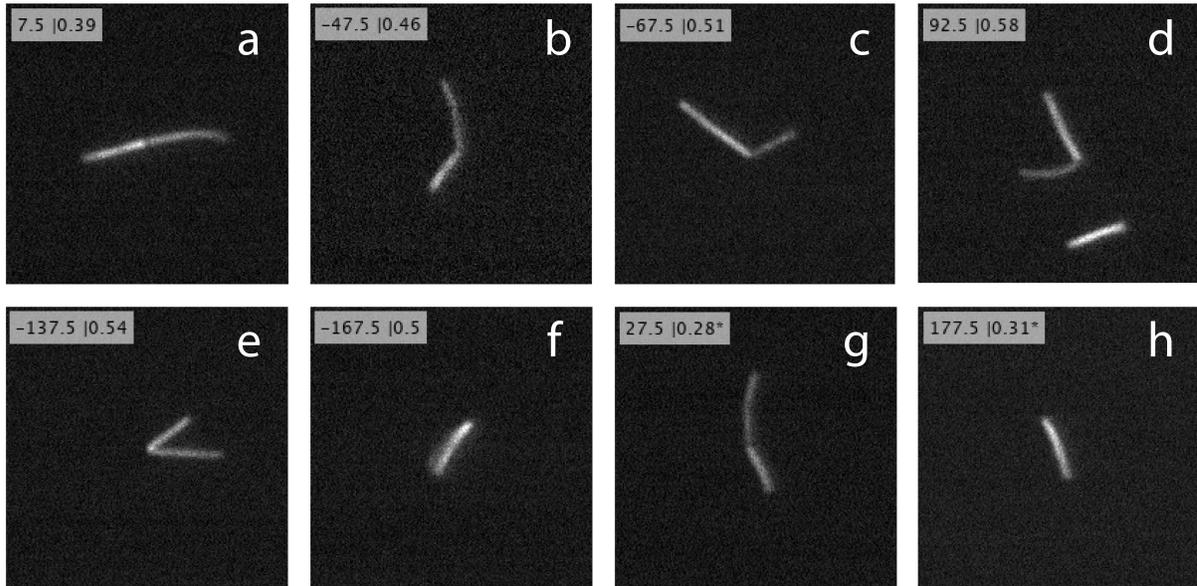
Figure 3.4: Real nunchuck images analyzed by NunchuckNet, displaying the determined bend angle and associated confidence score in the top left corner of each panel. The angles are denoted as the center of the classified bin. In panels (g) and (h), asterisks highlight confidence scores significantly below the mean, indicative of lower prediction certainty. Notably, panel (h) exemplifies a nunchuck with a bend angle exceeding 170°, a range where NunchuckNet's confidence is typically reduced.

and biases. The architectures and training protocols of NN2, NN3, and NN4 remained identical with those for NN1. Enhancements were only applied to the computer-generated training images, since the quality of the training set is paramount in improving the performance of any machine learning model. Our objective with subsequent networks has been to use training images that mirror the intricacies of real nunchuck images more closely.

NN2 was the first step in this progression. Adjustments included the length ratio of the nunchuck arms, a higher threshold for minimal arm brightness, and an emulation of contrast adjustments akin to those in data pre-processing. To better simulate the blurriness observed in experimental images, the diffusion parameter was increased. The introduction of Gaussian noise and the probabilities of arm kinks and brightness varia-

tions further enriched the dataset, while the inclusion of multiple background nanotubes added a layer of complexity previously unaccounted for.

NN3 expanded the potential location range for the nunchuck vertex, allowing it to be drawn within a 100 px × 100 px square instead of the previous 40 px × 40 px square. With NN4, the focus was on emulating the pre-processing steps that real images undergo before analysis. This version replicated the cropping procedure and incorporated more sophisticated noise models, including both Poisson and Gaussian, to closely approximate the statistical noise characteristics of real microscopy images.

To assess and compare the performance of NN1 through NN4, we used each to assess a set of 7,200 computer-generated nunchuck images uniformly distributed across 72 bend angle bins, which were synthesized using the advanced training image generation algorithm developed for NN4. This algorithm was selected because it generated images that most resemble real nunchucks. The results are depicted in Figure 3.5 as heat maps within confusion matrices. These heat maps visually represent the level of agreement between the neural networks' measured bend angles and the true angles. The accuracy of each neural network, annotated in the titles of the heat maps, reveals a clear trend: as the training sets are enriched with more realistic features, the neural networks demonstrate improved accuracy. This is illustrated by the increasingly tight clustering along the diagonal of the heat maps for each successive neural network version. While NN1 starts with an accuracy of 32.5%, subsequent networks show a measurable improvement, with NN2 at 36.8%, NN3 at 45.2%, and NN4 achieving 44.5%, on par with NN3. These metrics underscore the significant impact that enhanced training image quality has on the performance of neural networks.
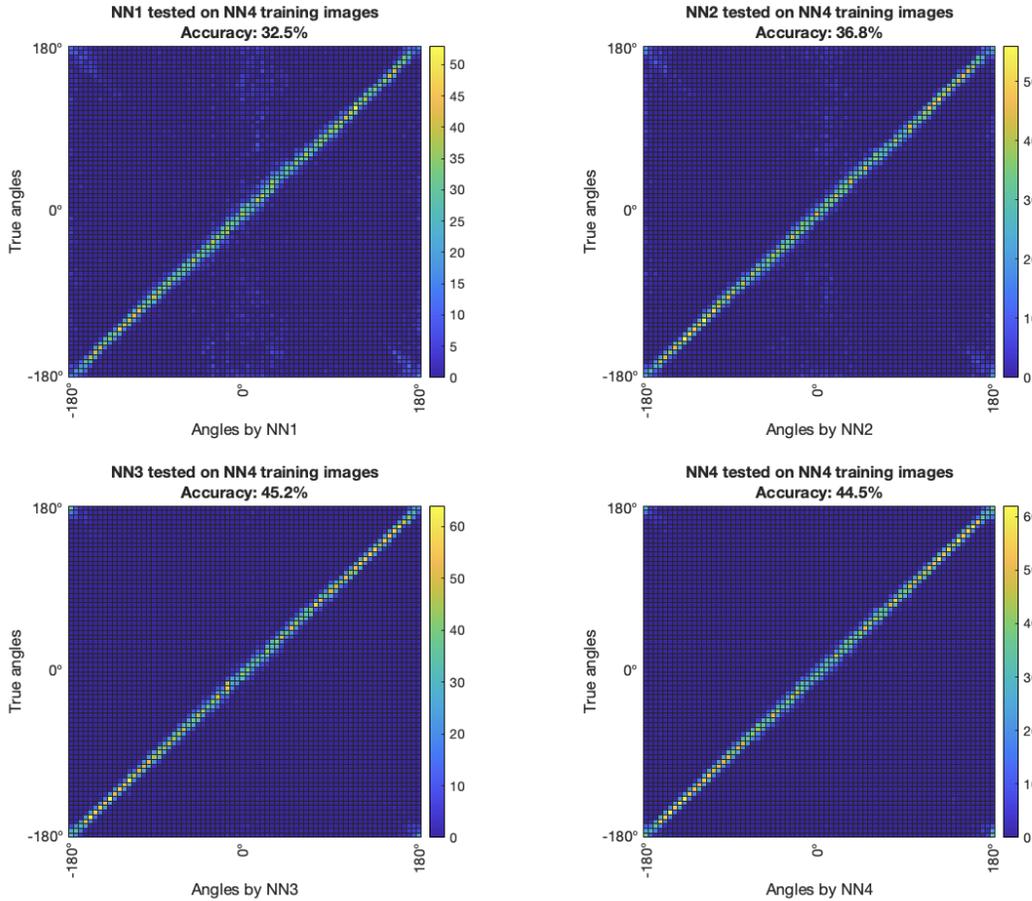
Figure 3.5:   Confusion matrix heat maps comparing the actual bend angles with the predictions made by four neural networks, NN1 through NN4, on a set of 7,200 algorithmically generated nunchuck images. These images were evenly distributed across 72 bend angle bins and were produced using the NN4 image generation algorithm for enhanced realism. Each heat map corresponds to one neural network's performance, with the accuracy score indicated in the title. The intensity along the diagonal reflects the number of correct predictions, with NN3 and NN4 exhibiting the highest accuracy, demonstrating the benefits of progressively refined training images on network performance.

3.1.5 Estimating the Uncertainty and Bias

In order to establish the reliability of the neural network measurements and align with standard practices in machine learning, we have evaluated uncertainties and biases in bend angle measurements from neural networks NN1 to NN4, as depicted in Figure 4.2. Each neural network was tasked with analyzing at least 720,000 computer-generated images, with the distribution of these images spanning uniformly across all 72 angle bins. Ideally, a neural network without any measurement bias would produce a histogram with uniform bin counts, *i.e.,* a flat histogram, signifying equal frequency across all angles. However, the observed variations in the histograms suggest inherent biases in the neural network measurements. The consistency of these biases for each neural network, as evidenced by similar histogram shapes in repeated tests with different sets of generated images, underscores their systematic nature. This pattern of bias indicates that certain bend angle bins are consistently overrepresented or underrepresented in neural network measurements. For all four versions of the neural network, we observe a consistent overestimation in the final bin and underestimation in the penultimate bin. This tendency likely arises from the networks' difficulty in differentiating highly bent nunchucks from nanotubes or from fully open nunchucks, a challenge previously noted. Although these biases in extreme bends do not severely impact the overall data analysis, owing to the rarity of completely closed nunchucks, they do inform us about potential improvements for future iterations of neural network training.

An intriguing observation from these results is the apparent trade-off between general accuracy and specific biases in the last two bins, which is a known phenomenon in machine learning. NN4, for example, demonstrates the highest general accuracy but also exhibits the most pronounced biases in the extreme bends. Conversely, NN3, while slightly less accurate across most bins, shows a reduced bias in these extreme angles. This
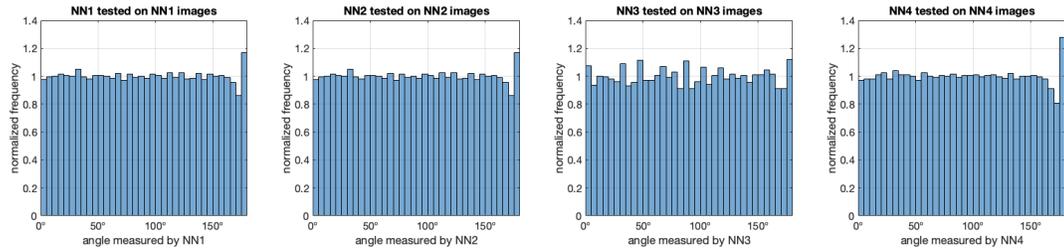
Figure 3.6: Histograms depicting the normalized frequency of bend angle predictions from neural networks NN1 to NN4. Each network was tested against at least 720,000 computer-generated images, uniformly distributed across 72 bend angle bins using NN4's algorithm. The histograms are normalized to show the expected frequency for a perfectly accurate network, with a flat distribution indicating no bias. The visible deviations from flatness in the histograms indicate consistent biases within the neural networks, with the last bins showing a notable overestimation and the penultimate bins an underestimation. This demonstrates a trade-off between overall accuracy and biases in extreme angle bins across the different network iterations.

phenomenon reflects an intrinsic aspect of the training process, where neural networks reach an equilibrium in minimizing error across all bins, with some networks tending to accumulate errors at the extremes.

3.1.6   Prospects for Improvement

Exploring branched neural networks

A persistent challenge with conventional neural networks is their "black box" nature, particularly when errors occur without clear indication of their root causes. One promising avenue for improvement lies in the deployment of branched neural networks. Such networks have the potential to mitigate issues with highly bent nunchuck images, cases where current models struggle to differentiate between a highly bent nunchuck and a nanotube or a straight nunchuck with a very short dim arm. By incorporating additional data channels that contain specific information provided by the user, such as the lengths of the nunchuck's arms, branched neural networks can provide additional context that helps the neural network make unambiguous decisions. Preliminary trials with this ap-

proach have shown promise, suggesting that a branched architecture could resolve some of the current limitations.

Advancing beyond AlexNet

Despite these innovations, there remains a ceiling to the capabilities of the current neural network model, AlexNet, which we suspect has been reached. This plateau is evidenced by the marginal change in accuracy seen from NN3 to NN4, despite the introduction of more realistic training images. To surpass this limitation, a shift towards larger and more complex neural network architectures such as ResNet may be necessary. These advanced models, with their increased depth and capacity for handling intricate features, hold the key to higher accuracy in bend angle measurements. However, their success hinges on the quality of the training data. Our preliminary tests show that larger networks can succumb to overfitting, achieving high training accuracy but performing poorly on actual datasets, unless synthetic training images can be generated to resemble real nunchucks even more. Thus, our future direction must encompass both the adoption of more sophisticated neural network architectures and the generation of highly realistic synthetic training images. Only through this dual expansion can we hope to achieve a significant leap in the precision of nunchuck bend angle measurements.

Chapter 4

Nunchuck Calibration

4.1 Nunchuck Bend Angle Distributions

In order to establish a robust methodology for measuring nunchuck bend angles, we sampled bend angles at a frequency of 1 Hz across an extensive array of individual nunchucks with varying linker lengths and sequences. For each distinctive linker type, we documented the behavior of at least a dozen nunchucks, with representative examples encapsulated in Movies S1 to S3 in our previous work [40]. An illustrative selection of bend angle histograms from seventeen nunchucks, all with a 32 bp linker containing no known intrinsic bends, is showcased in Figure 4.1. The bend angle (θ) is defined within the bounds of 0° and 180°, and thus each histogram has been modeled to a truncated folded normal (TFnormal) distribution (discussed in the next section with great detail) characterized by a mean (μ) and standard deviation (σ). For each nunchuck shown in Figure 4.1, we determined the parameters of the TFnormal distribution by using a least-squares fitting approach, with a bootstrapping technique iterated 500 times to quantify uncertainties in μ and σ.

The scatter plot illustrating the distribution of standard deviation (σ) and mean (μ)
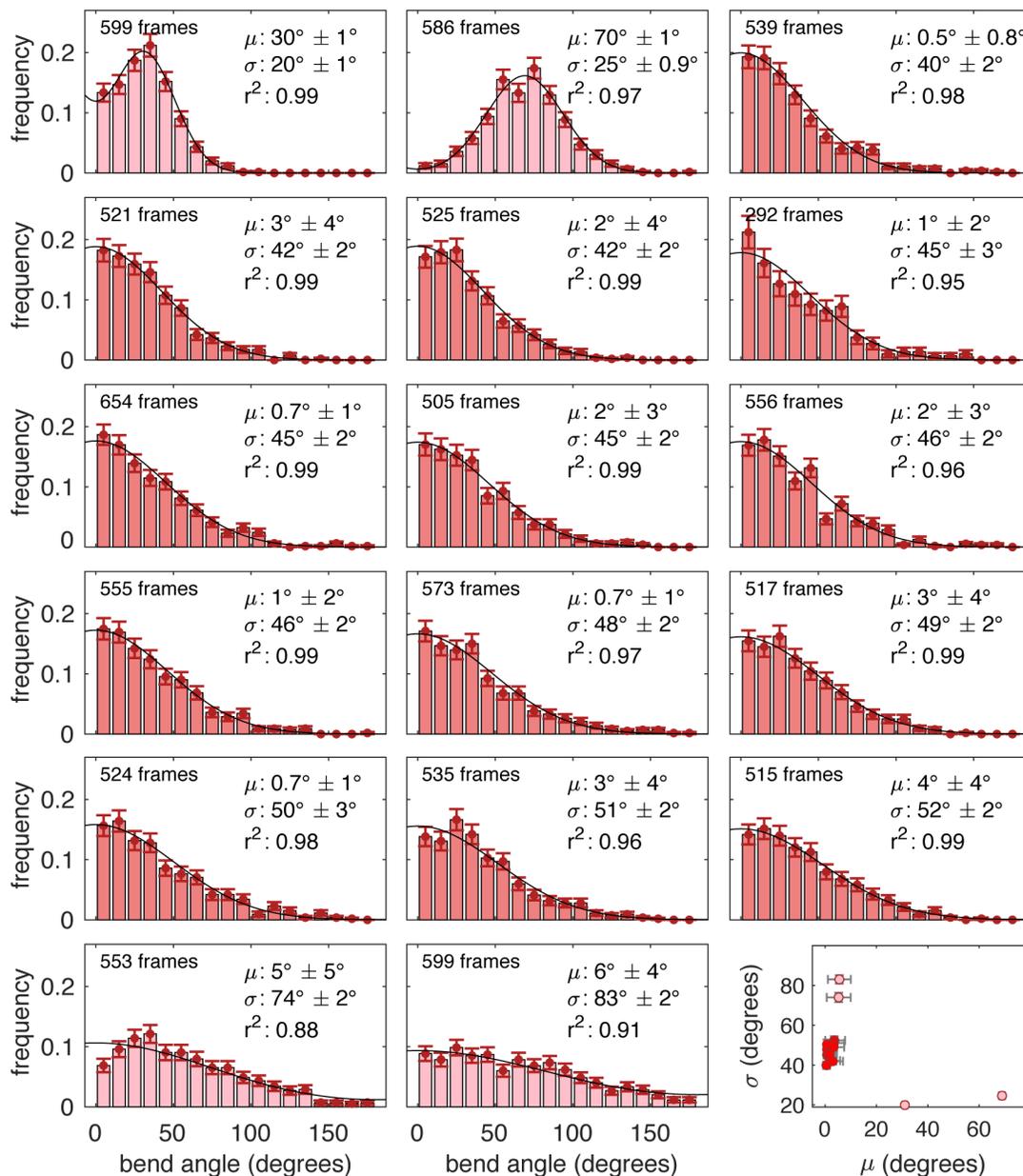
Figure 4.1:   Histograms representing the distribution of bend angles for nunchucks with 32 bp linkers, sampled at a frequency of 1 Hz. The displayed parameters, including the mean (μ) and standard deviation (σ), were deduced via least-squares fitting to a truncated folded normal distribution. The robustness of these parameters was evaluated through 500 bootstrap resamplings to estimate the uncertainties of μ and σ. The histograms are arranged in ascending order of σ, with the bottom right quadrant featuring a scatter plot that juxtaposes σ against μ for the entire dataset. Outliers, discerned by their lighter shade, have been excluded from the aggregate statistical analysis to ensure the integrity of the dataset.

of bend angle distributions (see Figure 4.1, bottom right) highlights a crucial aspect of nunchuck datasets. A significant subset, about 25%, diverge markedly in either μ or σ — surpassing two standard deviations — or align poorly with the truncated folded normal distribution, as evidenced by an $R^2 \leq 0.85$. As will be discussed thoroughly in upcoming sections, these vriations are not ascribed to environmental influences or imaging discrepancies, as they were often recorded at the same time and within the same visual scope as the majority, nor are they caused by structural defects such as incomplete ligation, or rogue strands interfering with nunchuck movements, as shown by a comprehensive array of tests. For the purpose of calibrating the nunchuck as a tool for describing linker stiffness and bend, we exclude these outliers while calculating set statistics, and focus only on clustered, regular distributions.

4.2   The Truncated Folded Normal Distribution

The nunchuck bend angle, denoted as θ, follows a Gaussian distribution. Observations, however, are only made within the interval $[0°, 180°]$. Consequently, bend angle data should be modeled by a distribution that results from folding a Gaussian over the specified interval. Any potential distribution existing beyond $[-360°, 360°]$ is considered negligible and is truncated, given that $\theta > 360°$ is exceedingly improbable. This adjusted distribution is thus termed the Truncated Folded Normal, or TFnormal for short.

The normal distribution is given by:

$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(-\theta-\mu)^2}{2\sigma^2}} \tag{4.1}$$

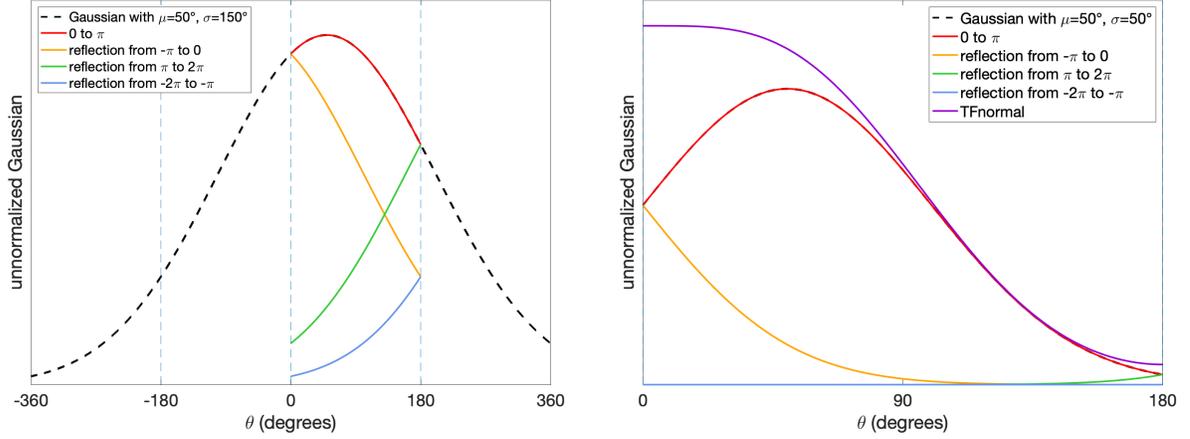where μ is the mean and σ is the standard deviation of the bend angles. Similarly, the

Figure 4.2:   Left: The truncation and folding effects on a Gaussian distribution (black dashed line) with mean $\mu = 50°$ and standard deviation $\sigma = 150°$ are shown. The red, orange, green, and blue curves represent the distribution from four different regions folded onto the interval $[0°, 180°]$. Distributions outside the range $[-360°, 360°]$ are truncated due to the extreme unlikelihood of observing angles greater than $360°$ in real nunchucks, typically characterized by $\sigma \sim 50°$. Right: The TFnormal distribution curve, depicted in purple, with parameters $\mu = 50°$ and $\sigma = 50°$, is representative of typical nunchuck behavior. This TFnormal curve is the summation of the folded red, orange, green, and blue curves.

TFnormal distribution is defined on $[0°, 180°]$ and is given by:

$$\frac{1}{\sqrt{2\pi\sigma^2}} \left(e^{-\frac{(-\theta-\mu)^2}{2\sigma^2}} + e^{-\frac{(2\pi-\theta-\mu)^2}{2\sigma^2}} + e^{-\frac{(\theta-\mu)^2}{2\sigma^2}} + e^{-\frac{(-2\pi+\theta-\mu)^2}{2\sigma^2}} \right) \qquad (4.2)$$

Unfortunately for nunchuck studies, The indistinguishability of certain TFnormal distributions presents a challenge in parameter estimation. For instance, TFnormal distributions with parameters $\mu = 0°$, $\sigma = 50°$ and those with $\mu = 20°$, $\sigma = 45°$ look nearly identical, as depicted by Figure 4.3. Without prior knowledge or assumptions about μ or σ, discerning the true parameters is not possible. Generally, estimating μ becomes problematic when $\mu < \sigma$ due to the folding-induced plateau at $\theta = 0°$, which obscures the distribution's peak. Therefore, while the TFnormal fitting approach is essential to nunchuck bend angle analysis, it encounters challenges when the mean (μ) is both un-

49
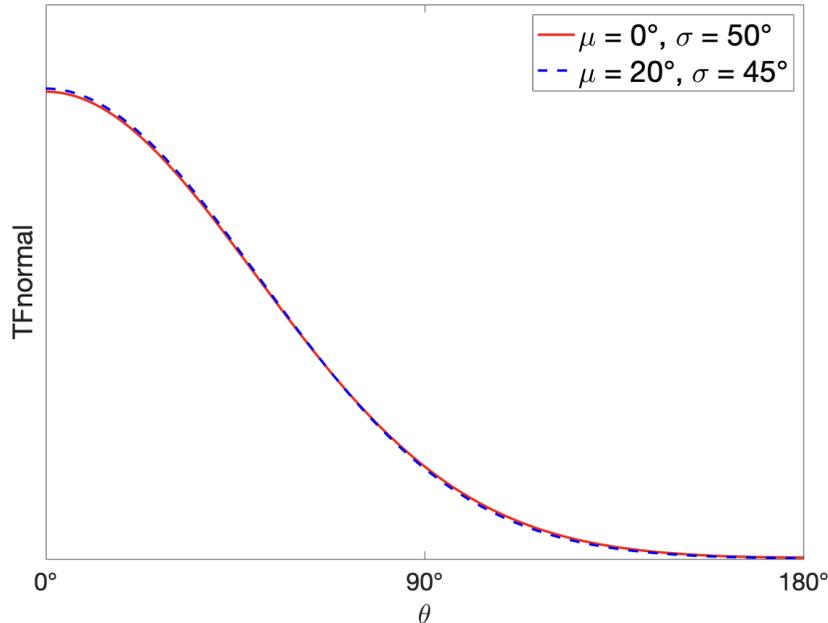
Figure 4.3:   Unnormalized TFnormal curves showcasing the difficulty in distinguishing between distributions due to the overlap of curves with parameters $\mu = 0°$, $\sigma = 50°$ and $\mu = 20°$, $\sigma = 45°$. This overlap illustrates the challenges in parameter estimation for TFnormal distributions without prior assumptions, particularly when μ is small and the plateau around $\theta = 0°$ masks the distribution's peak.

known and less than the standard deviation σ, resulting in ambiguous fits. To simplify matters when calibrating nunchucks for linker stiffness and bend, we impose a restriction that μ must be less than 10° for fitting, unless unrestricted fits demonstrate significantly better R^2 values.

To systematically characterize how this TFnormal fitting ambiguity affects the analysis of nunchuck data, we assess the indistinguishability between any two TFnormal distributions. This is done by computing their root-mean-square deviation (rmsd), which we define as the square root of the aggregated differences of the distributions, assessed at angles from 2.5° to 177.5° in 5° increments. These increments were selected as they correspond to bend angle values recorded by neural networks. For illustrative purposes, we consider the rmsd heat map between TFnormal$(0°, 50°)$ and another TFnormal distribution with specific μ and σ values plotted along the x and y axes, respectively, displayed
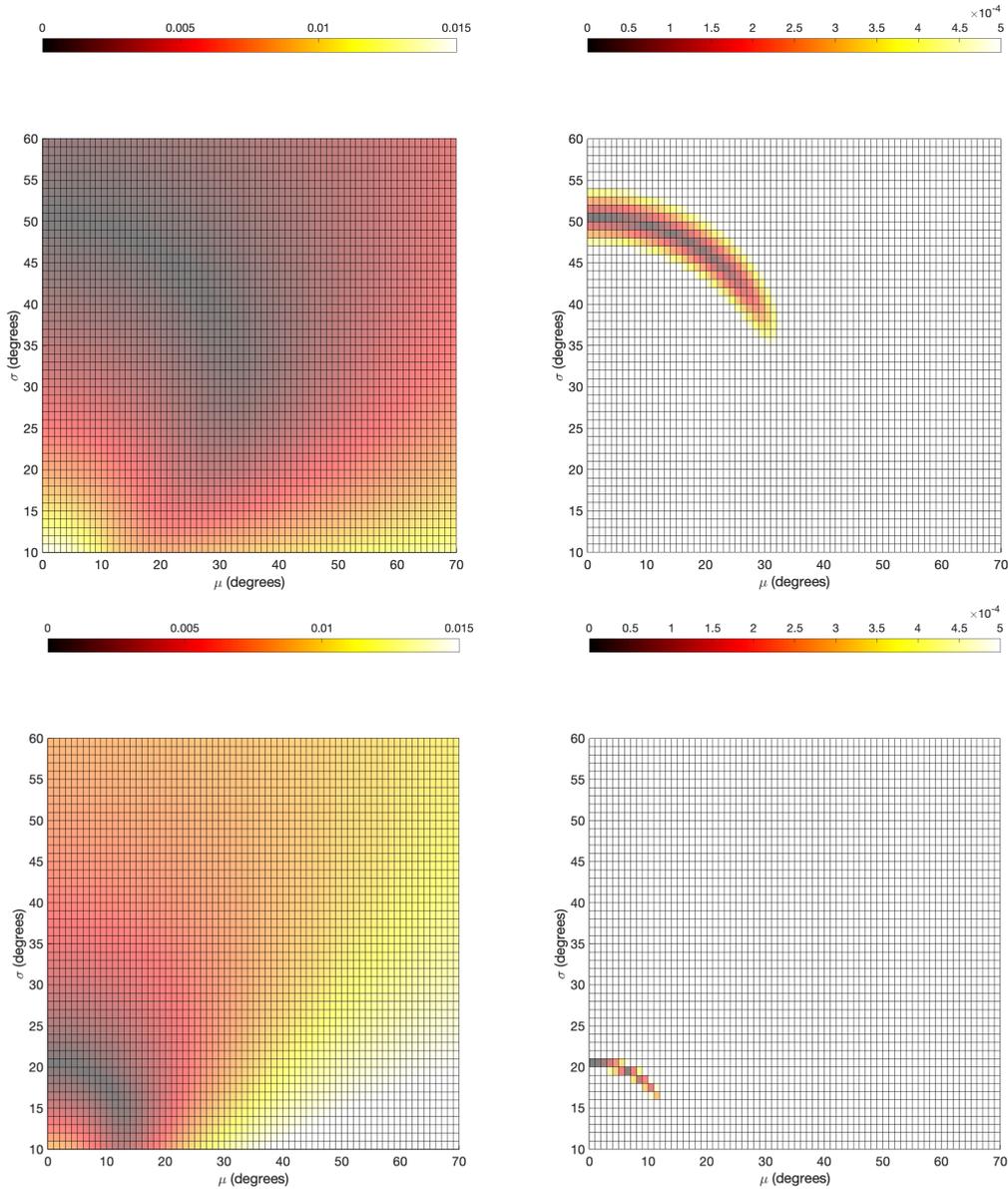
Figure 4.4: Top: Comparative heat maps displaying the root-mean-square deviation (rmsd) between TFnormal$(0°, 50°)$ and TFnormal with varying μ and σ values. The heat maps are juxtaposed to exhibit both the overall rmsd distribution across the full range of μ and σ (left) and to emphasize a narrowed region where rmsd is markedly low (right), indicative of the fitting ambiguity. Bottom: Heat maps of rmsd contrasting TFnormal$(0°, 20°)$ with a TFnormal variant with varying μ and σ values. The bottom maps utilize the same color scale as the top maps to maintain consistency. The area illustrating indistinguishability under a stringent rmsd criterion (less than 0.0005) is significantly diminished in the right panel, as opposed to the broader area observed in the TFnormal$(0°, 50°)$ scenario.

on two distinct color scales (Figure 4.4 top). Notably, the heat map on the right reveals a slender, arc-shaped zone with a minimal rmsd (less than 0.0005, transitioning from dark red to yellow), suggesting these TFnormal distributions are virtually indistinguishable from one another. In contrast, Figure 4.4 bottom illustrates a significantly reduced ambiguity region for TFnormal($0°, 20°$) when compared with TFnormal($0°, 50°$). This is due to the more constricted distribution of the $\sigma = 20°$ case, which consequently enhances the confidence in the precision of its peak location (μ). In summary, the ambiguity region of a TFnormal distribution with a small μ typically manifests as an arc-shaped area in the μ - σ space, but the extent and location of this region are contingent on the specific parameters of the TFnormal distribution.

4.3   Nunchuck Angular Variance Scales with Linker Length

Having introduced the TFnormal distribution and demonstrated some examples of nunchuck bend angle distributions, we investigate the relationship between angular variance, denoted as σ, and linker length. In order to calibrate the nunchuck as a tool for measuring linker bending, it is essential to show that nunchuck flexibility is solely determined by the flexibility of the linker, and to extract this exact relation from a linear fit to data. The bend angle, θ, observed in a nunchuck, is defined as the angle between the tangent vectors at the ends of the linker. According to the worm-like chain (WLC) model, the average of the cosine of this bend angle is related to the 3-dimensional persistence length of the linker, L_p, as

$$\langle \cos\theta \rangle = e^{-\lambda/L_p} \tag{4.3}$$

where λ is the contour length separating the tangents. In this case, the 3-dimensional persistence length L_p is used rather than the 2-dimensional one, because the nunchuck's arms, given their dimensional scale compared to the gap between PEG-passivated glass surfaces, can freely adopt any 3-dimensional conformation without energetic penalty, facilitated by the freedom of rotation to maintain parallelism to the plane of confinement. Moreover, room temperature thermal energy can introduce significant twist in the linker, further helping it to achieve all possible bent states.

The distribution of bend angles θ is modeled by a normal distribution with variance σ^2. For such cases, it is easy to show that (see Supplementary Note S14 of our previous work[40])

$$\langle \cos \theta \rangle = e^{-\sigma^2/2} \ , \tag{4.4}$$

which, when combined with the WLC model, implies a direct proportionality of angular variance with the effective linker length, λ:

$$\sigma^2 = \frac{2\lambda}{L_p} \tag{4.5}$$

where we anticipate that the effective linker length, denoted as λ, may diverge from the defined (nominal) linker length, l, expressed in base pairs. This discrepancy is due to design considerations, such as the junctions where linkers integrate into origami seeds, potentially introducing additional flexibility. Consequently, the effective linker length could be conceptualized as the sum of the nominal linker length and a constant offset. This offset is an inherent property of the nunchuck and has to be determined empirically from the bending data of the nunchucks.

We acquired image sequences for nunchucks with various linker lengths: 27, 32, 37, 42, 47, 52, and 58 bp respectively. All linker sequences can be found in the Supplementary

Note S4 of our previous publication [40]. These numbers were selected to produce integer

or half-integer number of helical turns so that we can see how nunchuck bending is affected

by the phasing of the nunchuck arms with respect to linkers. Bend angles were extracted

from image sequences using the neural network, and TFnormal fits were used to extract

the mean bend angle μ and standard deviation σ of each nunchuck.

The results align with the model that σ^2 scales with λ. When considering linkers

with seeds on opposite sides of the helical axis (the "out-of-phase" configuration), the

angular variance grows linearly with increasing linker length (Fig. 4.5a, black circles).

This linear correlation provides an estimated persistence length of $L_p = 131 \pm 33$ bp,

consistent with previous dsDNA measurements under similar ionic conditions[51]. In

contrast, when linker lengths are integer helical turns as to place the seeds on the same

side of the helical axis (Fig. 4.5a, red squares), the angular variance follows the predicted

trend only for shorter linkers where seed collisions are infrequent due to the relative

stiffness of the linker. At greater lengths, the variance is notably diminished, presumably

due to steric hindrance between the seeds. An intermediate case is also observed where

the angular variance for a 55 bp linker (Fig. 4.5a, blue triangle) aligns with expectations,

situated between the variance for linkers of 52 bp and 58 bp in length.

The presence of a non-zero intercept in the linear fit is expected as discussed above.

This additional flexibility is quantified as an effective length, l_0, by which the effective

linker length, λ, exceeds the defined linker length, l, such that

$$\lambda = l + l_0 \tag{4.6}$$

The value of $l_0 = 10 \pm 11$ bp extracted from the fit implies that the bending of the linker

extends into the adjoining seeds, approximately spanning half of a helical turn. This

could also be interpreted as the inherent flexibility at the point where the linker and

seed meet. We have thus calibrated the nunchuck for use in measuring intrinsic linker flexibility, both by confirming that the nunchuck flexibility scales linearly with linker flexibility, and by extracting l_0 so that the effective linker length can be calculated for any future linker design.

4.4   Nunchuck Mean Bend Angle Reflects A_6-tract Static Bend

We next seek to test the nunchuck as a tool for measuring the intrinsic bend angle of its linker. For this purpose, we have selected phased A_6-tracts, which consists of a series of 6 adenine bases appropriately spaced so that the intrinsic bends of each run of adenine bases add up in a co-planar fashion. We measured the bending for nunchucks with various linker lengths containing various number of A_6-tracts, and used TFnormal fits to obtain the mean bend angle μ for each linker design. By applying a linear fit to the data (Fig. 4.5b), assuming a zero intercept ($\mu = N\alpha$), where N represents the number of A_6-tracts within the linker, we derived an angle α of $17° \pm 1°$. This finding aligns closely with previous studies and exhibits a comparable level of precision (as reviewed by Stellwagen $et\ al.$[52]).

Our study also revealed that nunchucks with A-tract-rich linkers exhibited significantly lower angular variances compared to those without these tracts (as shown in Figure 4.5a). For instance, linkers with 37 bp without any A_6-tracts demonstrated a standard deviation of $49° \pm 4°$. In contrast, linkers of the same length but containing three A_6-tracts showed a reduced standard deviation of $41° \pm 3°$. To discern whether this reduction in variance stemmed from the A_6-tracts themselves or other factors such as steric hindrance, we examined 37 bp linkers with three A_6-tracts arranged in a special
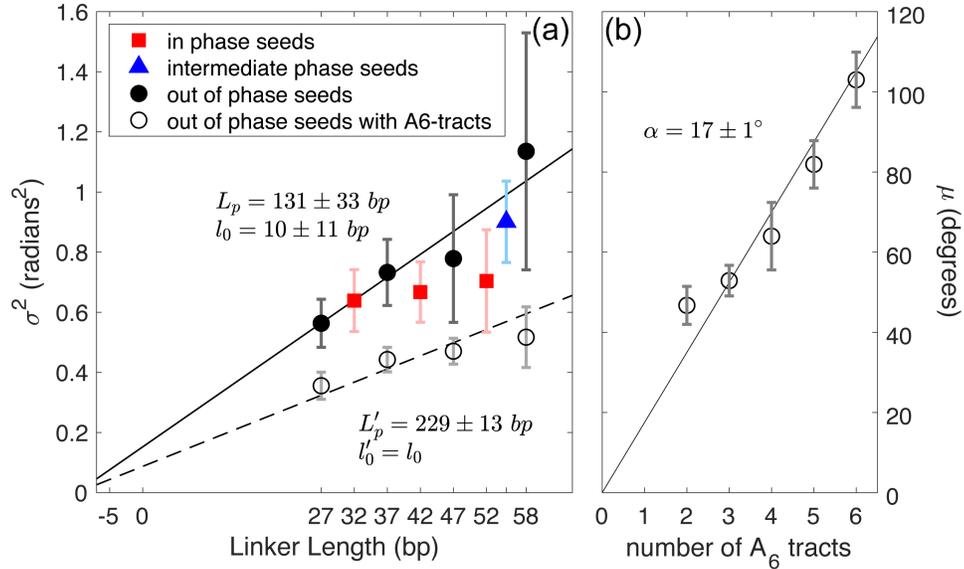
Figure 4.5:   (a) Filled markers represent the weighted average of angular variances from a sample size of at least 12 individual nunchucks, each containing no intrinsic bend, plotted against the defined linker length in based pairs. The trend of increasing angular variance with linker length aligns with the expectations derived from the WLC model, and is specifically noted for cases where seeds are positioned on alternating sides of the helical axis (indicated by black circles), the so called "out-of-phase" positioning. Conversely, linker lengths that arrange seeds on the same helical side (indicated by red squares) exhibit the predicted linear relationship only when short enough to prevent seed collisions. For longer linkers, the trend deviates due to potential steric hindrances. The angular variance for a linker measuring 55 bp (marked by blue triangles) appropriately falls between the variance values for 52 bp and 58 bp linkers. A linear fit applied solely to the data represented by black circles provides an estimate for the persistence length of the dsDNA linker, L_p, and the distance, $l_0/2$, at which the linker orientation is anchored into a seed. The error bars indicate the weighted standard deviation of the angular variances. Open markers illustrate the angular variance relative to linker length for linkers packed with phased A_6-tracts. Fitting this data to Eq. 4.5 with an established $l_0' \equiv 10$ bp (derived from l_0), we determine the persistence length for phased A_6-tracts to be $L_p' = 229 \pm 13$ bp. (b) The graph plots the average bend angle, μ, versus the number of phased A_6-tracts, N, in the linker. When $N = 3$, μ represents the weighted mean calculated from data across 27, 37, and 47 bp linkers. For $N = 2, 4, 5, 6$, the data is sourced from single linkers of lengths 37, 37, 47, and 58 bp, respectively. The linear relationship between μ and N, expressed as $\mu = N\alpha$, allows for the determination of an intrinsic bend angle of $\alpha = 17 \pm 1°$ per A_6-tract.

phasing to produce no net bend. These linkers exhibited a mean bend angle of $0° \pm 1°$ and a standard deviation of $43° \pm 5°$, similar to that of the intrinsically bent 37 bp linkers with three A_6-tracts.

While it has been hypothesized that A_6-tracts are inherently stiffer than generic dsDNA (as discussed in Stellwagen *et al.*[52, 53, 54]), experimental validation of this stiffness had not been previously conducted to our knowledge. By comparing linkers of varying lengths and similar (highest possible) densities of phased A_6-tracts (27 bp with $N = 3$, 37 bp with $N = 4$, 47 bp with $N = 5$, and 58 bp with $N = 6$), and aligning this data with Eq. 4.5 under the condition $l'_0 = l_0 = 10$ bp, we determined an increased persistence length $L'_p = 229 \pm 13$ bp, which suggests an enhancement of approximately $(80 \pm 30)\%$ relative to generic dsDNA.

Chapter 5

Variability Among Structurally Identical Nunchucks

Despite the success with nunchuck calibration so far, as mentioned in the previous section, we have encountered a perplexing phenomenon with every dataset: identically prepared nunchuck molecules often exhibited highly distinct bend angle distributions, which up to this point has led to us to exclude the outliers from statistical analyses through clustering algorithms. As an example, the variability in individual nunchuck behavior was notable among nunchucks containing A_6-tracts within their linkers, illustrated by Figure 5.1, wherein each set with a specified linker presented at least one pair with distributions that diverged significantly from each other, as validated by the stringent criteria of the Kolmogorov-Smirnov test ($p \ll 1/N_{\text{pairs}}$). One potential explanation is that linkers enter short-lived states which affects its distribution of bend angles, such as the one shown in Figure 5.2, where the nunchuck oscillated between a highly bent state and a less bent or unbent state, underscoring the occurrence of such transient states.

To investigate the underlying causes of the observed variability in behavior among the nunchuck populations, and to assess the likelihood that such behavior could emerge by
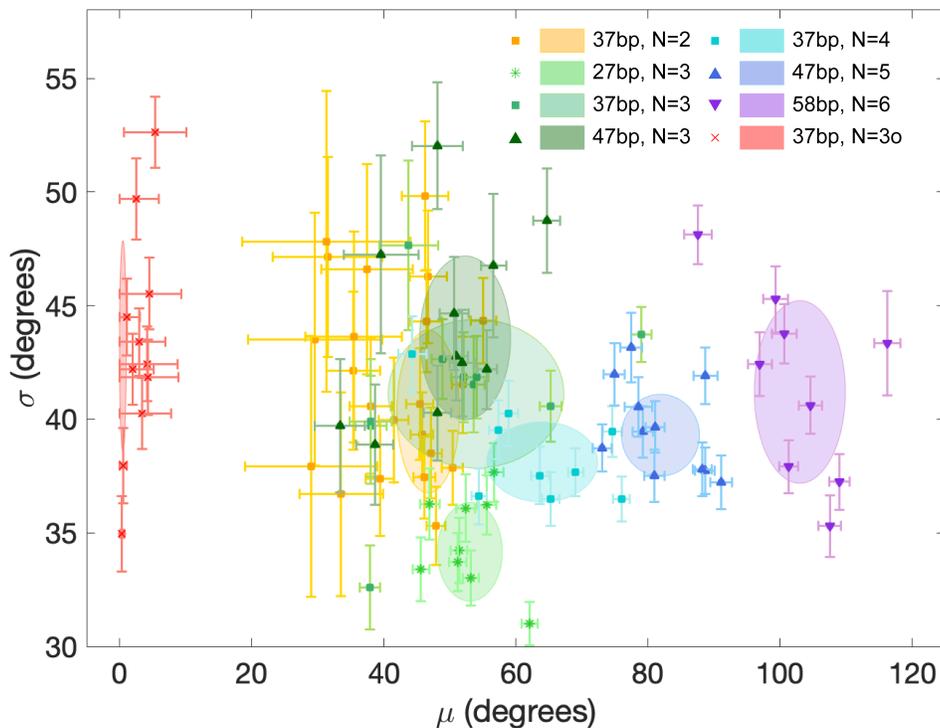
Figure 5.1: This scatter plot graphically represents the standard deviation (σ) and mean bend angle (μ) of nunchucks that incorporate A_6-tracts in their linkers. The data points are color-coded to differentiate between linkers containing varying numbers of A_6-tracts, with the hues and marker styles further distinguishing the linkers according to their lengths. Overlaid on this scatter plot are ellipses, each matching the color and shade of their respective data sets. These ellipses are centered around the weighted average values of μ and σ for each set. The horizontal and vertical spans of the ellipses represent the weighted standard deviations of μ and σ, respectively, visualizing the distribution and variance within each group. The visualization facilitates a comparison among different linker configurations, allowing for an assessment of how the number and phasing of A_6-tracts affect the bending behavior of DNA linkers.
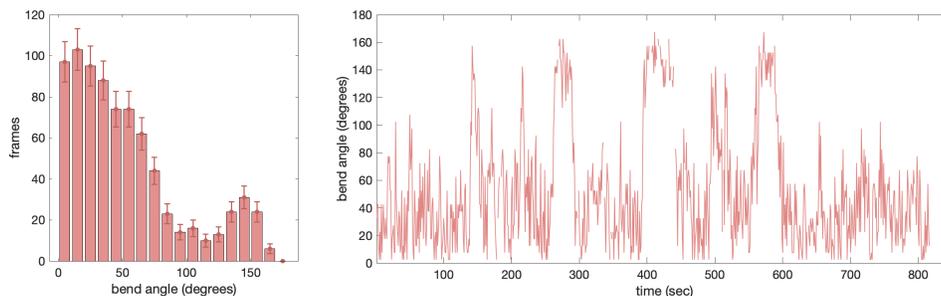
Figure 5.2: Left: A histogram depicting the distribution of bend angles for a nunchuck with a 58 bp linker that lacks a (known) intrinsic bend. The bimodal distribution is evidenced by two distinct peaks. Right: A time-resolved plot of bend angles, where each point represents the bend angle at a specific time increment. Gaps in the data indicate instances where the bend angles could not be confidently determined from images due to factors such as suboptimal imaging conditions or the nunchuck transiently exiting the observation field. These instances are identified by a low confidence score from the neural network analysis.

random chance due to large data sample sizes, we analyzed the bend angle distributions from an extensive dataset comprising 191 total nunchucks, each containing the identical 37-bp linker. Upon rigorous statistical evaluation, it was determined that merely 71 of these nunchucks (approximately 40% of the sample) displayed bend angle distributions that conformed well to a TFnormal model. This conformity was ascertained with a 68% confidence level, established by comparing the sum of squared errors from the best TFnormal model against the 68% confidence interval of a corresponding Chi-squared distribution. This analysis provided a quantitative basis to refute the hypothesis that the outlier behavior was merely a statistical anomaly, thus prompting further investigation into alternative explanatory factors.

5.1   Eliminating Confounding Factors

We systematically examined a spectrum of factors that could potentially account for the divergent behavior observed in individual nunchuck assemblies. These factors included: (1) the immediate environment surrounding each nunchuck, (2) structural

imperfections within the linker, such as imperfectly ligated nicks, (3) the 300 bp unfolded single-stranded DNA within the scaffold strand, (4) the single-stranded DNA present in the fringe region at the proximal ends of the seeds, (5) single-stranded DNA contaminants within the experimental environment, and (6) the previously addressed issues related to the fitting ambiguity inherent in the TFnormal distribution. In the ensuing sections, we will discuss each hypothesized mechanism, elaborate on the experimental and analytical procedures employed to investigate these mechanisms, and present our findings, which collectively demonstrate that these factors cannot account for the variability in behavior manifested by the nunchucks.

5.1.1   Environmental Factors

Nunchucks were often observed to spontaneously enter an unambiguously bent state and remain bent for several seconds or more. Since data were taken with nunchucks mechanically confined to a small gap between a glass slide and a coverslip, we examined whether the nunchuck bend state transitions could be attributed to features of these PEG-coated glass surfaces. We note that "sticky" patches can exist on PEG-coated glass surfaces due to imperfect coating, and occasionally nunchucks were found stuck on the glass surfaces either permanently or for a few seconds. Such cases were easy to identify by eye, and "sticky" nunchuck videos were not analyzed for data. All bend angles reported in this dissertation came from nunchucks which could roam freely and were never observed being stuck. Hence, bend state switches were not caused by interactions between nunchucks and glass surfaces.

To test whether bend state switches could be attributed to other surface features such as, say, a local topology in the PEG-coated surface that prevented a passing-by nunchuck from being straight, we examined bend angle time series of nunchucks imaged
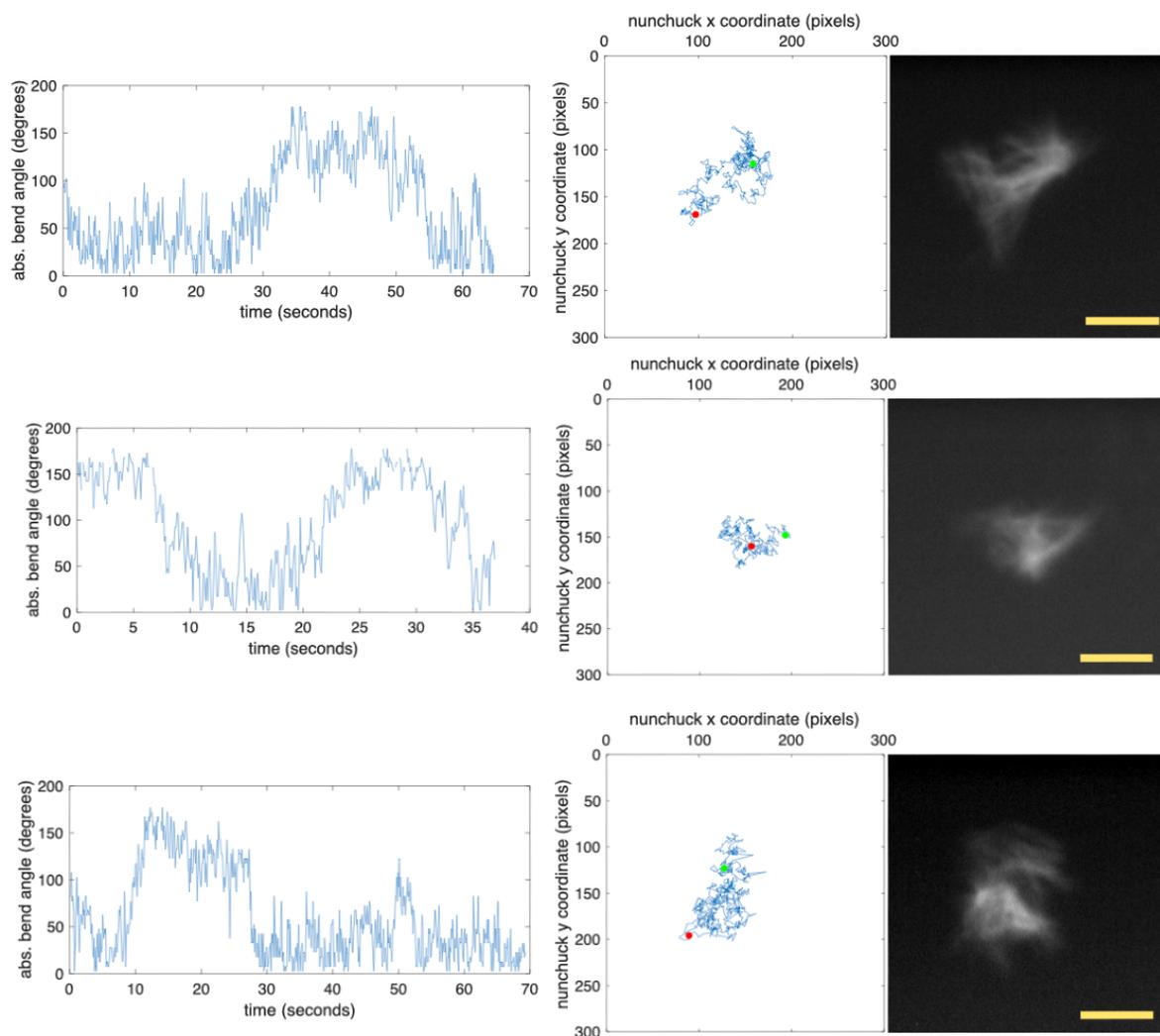
Figure 5.3:   From top to bottom: Examples of three different nunchucks with 37-bp linkers. For each example, left: the bend angle time series of this nunchuck; middle: the corresponding trajectory of the nunchuck's translational diffusion over the glass surface. The start point and end point of the trajectory are marked in red and green, respectively; right: brightness of the nunchuck averaged over all frames from the corresponding period. Scale bar: 5 μM. The middle and right figures are on the same scale.

at $\sim$10 Hz. Specifically, we looked at periods in which the bend angle unambiguously switched between a straight state and a bent state, and asked whether the switch was accompanied by the physical movement of the nunchuck into a different region on the glass surface. Figure 5.3 shows three such examples. In each case, both the trajectory of the nunchuck's center of mass (middle figures) and the averaged brightness over all frames (right figures) show that the bend state switch was not caused by the nunchuck entering a distinct region on the glass. While we do not show here every transition we studied, and we did not investigate every case of bend state switch, these three examples show that the presence of bent states cannot be attributed to environmental factors such as local landscape or stickiness.

5.1.2   Linker Nicks

To investigate whether variability in nunchuck behaviors could be ascribed to nicks within their linkers, we first determined the ligation success rate of the linkers. This was accomplished by treating ligated dimer seeds with exonucleases that digest unligated single and double-stranded DNA. A detailed protocol for this test can be found in Supplementary Note S10 of our previous work [40]. Agarose gel electrophoresis (Supplementary Note S10, Figure S9) and subsequent analyses revealed that approximately 70% of dimers were free from nicks after a standard ligation period, with an improvement to 80% after extended ligation.

Therefore, the ligation was not perfect, but we could not determine from this test whether the nicks have contributed to the variability in nunchuck behavior. Hence, we proceeded to compare the angular distributions of nunchucks equipped with three variants of a 37 bp linker of the same sequence: those with two ligatable nicks, those with one single ligatable nick, and those without any nicks. This comparison was essential to ascertain if

Figure 5.4:   The design and sequences of a 37 bp linker that contains no nick, in contrast to the default 37 bp linker used in this study, which contains two ligatable nicks (see Figure 2.5).

the presence of nicks could account for the variability in nunchuck behavior. The results of this comparative study are encapsulated in Supplementary Note S10, Table S6 of our previous work. The bend angle statistics for nunchucks, regardless of the number of nicks in their linkers, were remarkably consistent. The mean bend angles (μ) and standard deviations (σ) of the nunchucks were within the uncertainties of one another across all three sets. This consistency strongly suggested that the variability observed in nunchuck behaviors could not be attributed to the presence of nicks within the linkers.

5.1.3   Excess Scaffold Strand

The next possibility to consider in our quest to understand nunchuck behavior variability is the possible interaction attributed to the uncompacted 300 bp excess ssDNA in the scaffold. The DNA nunchuck design utilizes a 2,688 bp segment to fold the seed origami structure, leaving an overhang of approximately 300 bases from our P3024 scaffold strand. In some design iterations, these excess bases were compacted into a 6-helix bundle, a configuration detailed in Figure 5.5 below. Our investigations, as elucidated in Supplementary Note S12 of our previous work, revealed that the presence or absence of this compacted structure had a negligible impact on the bending behavior of nunchucks. Whether the 300 bp overhang was left loose or neatly packed into the 6-helix bundle, the
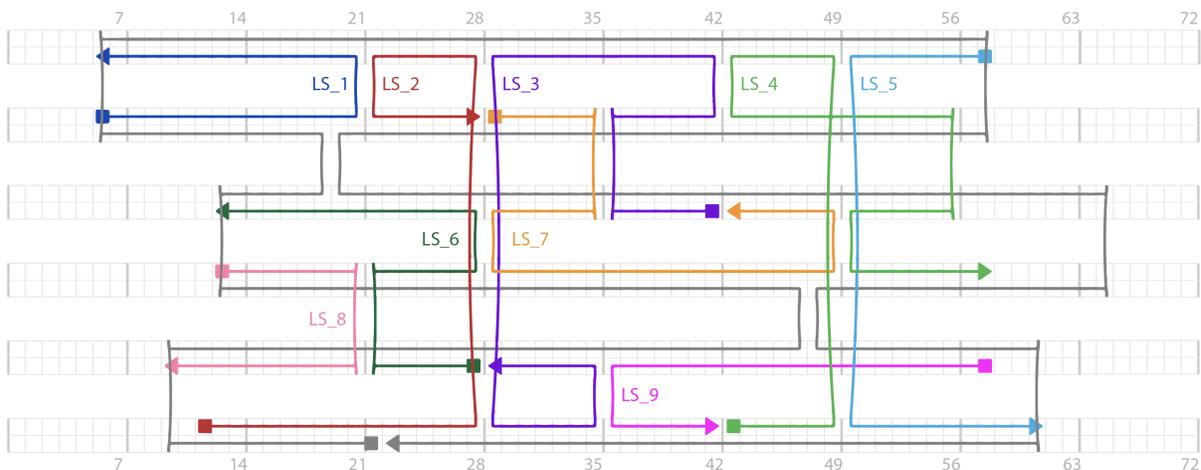
Figure 5.5:   Schematic representation of the seed scaffold with 316 out of the 336 excess bases configured into a 6-helix loop bundle. Staples are named "L(oop)S(taples)_index". The bundle, approximately 50 base pairs in length, translates to an estimated 17 nm, which spans about one-quarter of the seed's total length (224 bp or roughly 75 nm).

resulting nunchuck structures exhibited no significant variance in their bending properties. This finding allowed us to effectively rule out the excess scaffold strand, in both compacted and uncompacted forms, as a contributory factor to the variability in behavior observed in nunchuck assemblies. Furthermore, when we utilized a shorter scaffold variant, p2717, which is a subset of the p3024 sequence and consequently has a reduced overhang of extra bases that does not necessitate compaction, we observed that the bending characteristics of nunchucks constructed with p2717 were indistinguishable from those assembled with the p3024 scaffold.

5.1.4   Single-Stranded Fringes

We next identified the single-stranded regions at the ends of seed origamis — referred to as "fringes" — as potential contributors to variable nunchuck behavior. These fringes are designed to be left single-stranded on the proximal ends of the seeds to prevent
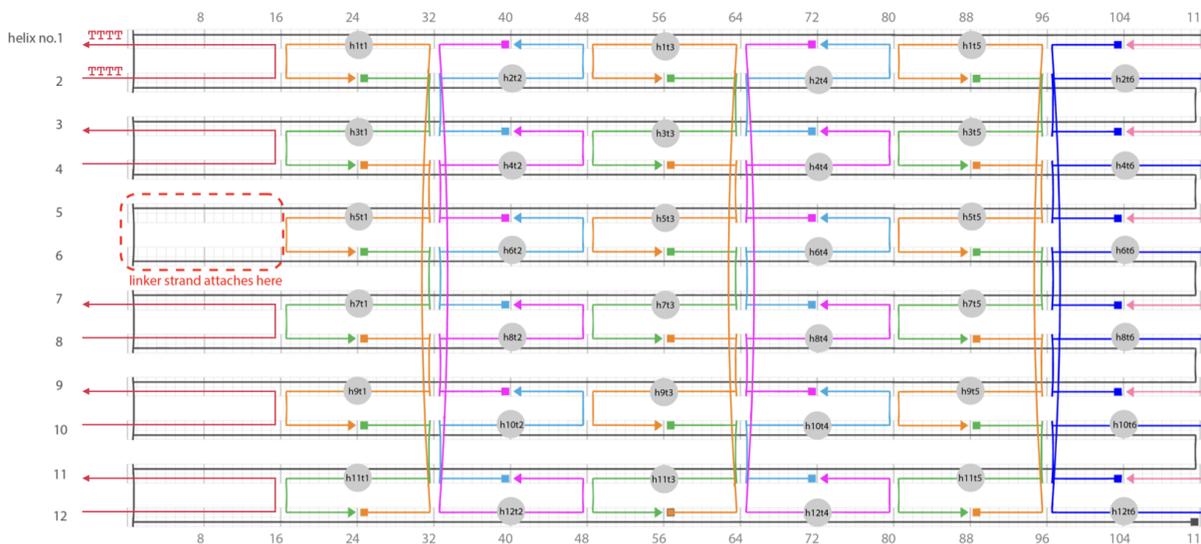
Figure 5.6:   An illustration of half a nunchuck seed with fringe blockers. The seed scaffold is shown in dark grey, staple strands are shown in various colors, and fringe blockers are shown in red. There are 5 fringe blocker strands corresponding to the 10 free helices which are not linker attachment sites.

stacking during dimer formation. Conversely, on the distal ends, they are hybridized to adapter strands. If these ssDNA fringes were instead hybridized by complementary strands on the proximal ends, they would hinder the proper formation of dimers due to stacking, as shown by our tests. However, the presence of such ssDNA fringes introduces the risk that they might bridge with free ssDNA in the environment or interact with each other, potentially influencing the bending behavior of the nunchuck.

To investigate whether these fringe regions have an effect on nunchuck bending, we designed "fringe blocker" strands (Figure 5.6) to preclude rogue strands from interacting with the fringes of scaffold DNA at the proximal ends of the seeds. They were annealed with the seeds, and added in equal concentration to the staple strands. Our data showed that the incorporation of fringe blockers did not significantly alter the bending distributions, nor did it appreciably reduce the occurrence of outliers. This outcome suggests that the fringe regions, even when unprotected by the fringe blockers, do not play a

decisive role in the variability of nunchuck bending behaviors.

5.1.5   ssDNA in the Surroundings

Next, we turned our attention to the possible role of rogue ssDNA strands within the sample environment. An excessive amount of ssDNA strands, if present, could potentially interact with the nunchuck structures, leading to discrepancies in observed bending behaviors. To investigate this, we employed mung bean nuclease (MBN), an enzyme known for its specificity in degrading ssDNA without affecting double-stranded DNA (dsDNA), thereby eliminating any extraneous ssDNA that might be interacting with the nunchucks. This targeted approach was intended to cleanse the nunchuck environment of any ssDNA that might confound our observations.

Before MBN was applied to nunchucks, we had to establish a protocol for MBN treatment that effectively removed ssDNA while preserving the integrity of nunchuck structures, as the MBN at too high a concentration can damage dsDNA too. Using a model system of Cy3-labeled REp and SEp tiles, we simulated nunchuck conditions, maintaining DNA concentrations comparable to nunchuck samples. The tiles were carefully annealed and treated with different MBN concentrations, followed by proteinase K incubation to prevent further damage to the DNA double helix. Fluorescence microscopy analysis revealed that nanotube formation, indicative of ssDNA presence, was significantly hindered with 1.5 units of MBN and completely prevented at 10 units in a 20 μL reaction. Controls without MBN displayed expected nanotube lengths. These experiments, conducted in neutral pH and with a zinc ion concentration of 1 mM, provided a reliable treatment regime to be applied to nunchuck assemblies, ensuring the removal of rogue ssDNA without compromising the nunchuck's structure. With the optimal conditions determined, we proceeded to apply the MBN treatment to actual nunchuck

samples. The nunchucks, featuring a 37 bp linker, were first ligated to ensure structural stability during subsequent enzymatic processing. Following ligation, the nunchucks were exposed to MBN under the established conditions, with a subsequent incubation with proteinase K to deactivate and remove the MBN. This step was critical, as any residual nuclease activity could potentially degrade the nunchucks themselves during the bending behavior assays. After the MBN treatment, we analyzed the bending distributions of the treated nunchucks. Despite the thorough elimination of ssDNA, which we expected to standardize bending behaviors, the variability within bend angle distributions remained. This confirmed that the presence of rogue ssDNA was not responsible for the observed variability in nunchuck bending.

5.1.6   TFnormal Ambiguity

The inherent fitting ambiguity in TFnormal distributions, as previously mentioned, can lead to a situation where the best fit for a nunchuck bend angle distribution fails to accurately represent the true distribution. Instead, it may yield entirely incorrect values for the parameters μ (mean) and σ (standard deviation). Specifically, bend angle distributions that appeared consistent with a TFnormal distribution having a mean bend angle of 0° could, due to the properties of the distribution, also fit a distribution with a non-zero mean, such as 30°. This inherent feature of TFnormal distributions raises questions about the statistical significance of the variability observed among the bending behaviors of individual nunchucks.

Figure 5.7 shows the best fit parameters μ and σ for 71 nunchucks with identical 37 bp linkers. This collection contains only nunchucks for which the bend angle distribution fits well to a TFnormal distribution, as judged by a confidence interval of at least 68%. Despite the good fit qualities, there is still considerable variability among
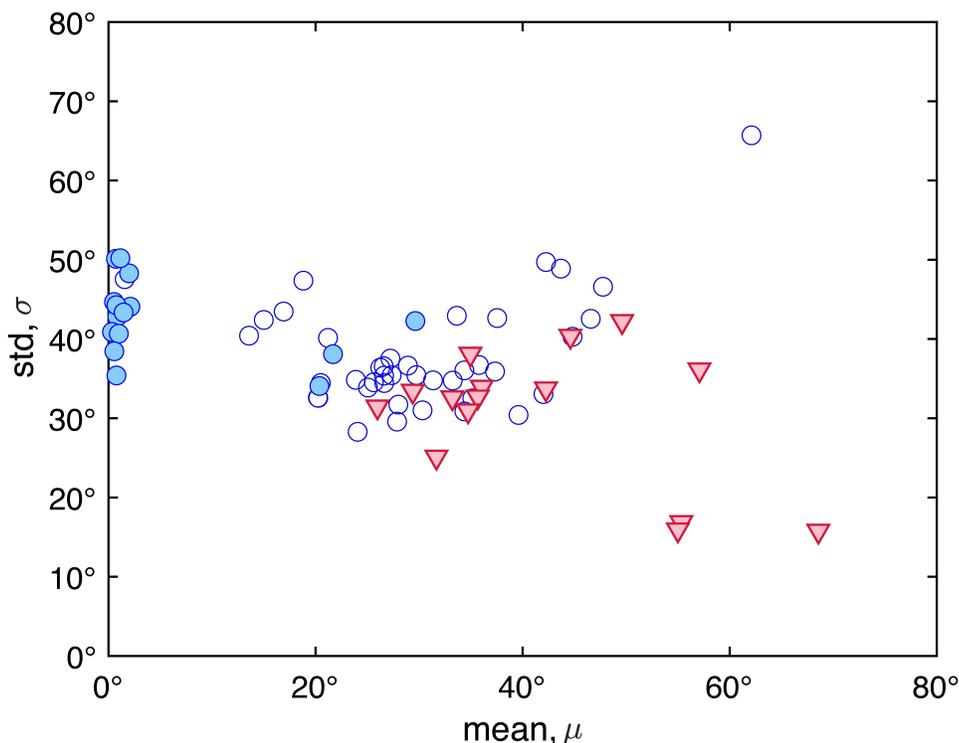
Figure 5.7:   Graphical representation of bend angle parameters from 71 nominally identical 37 bp nunchucks that fit well to a truncated folded normal distribution. Each marker indicates the bend angle distribution for a single nunchuck. Red filled triangles correspond to distributions which fit $\mu > 10°$ with at least 95% confidence, as determined by simulation. Blue filled circles correspond to distributions which fit $\mu < 10°$ with at least 68% confidence, as determined by simulation. All other distributions are plotted as hollow blue circles.

individual nunchucks. To test whether the effect of the TFnormal ambiguity can explain the nunchuck variability seen in this figure, we performed an elaborate simulation study. The aim was to understand whether the variability seen in our results could be a consequence of the fitting artifact associated with the TFnormal model. The simulation process involved generating a probability distribution of simulated best-fit parameters (μ' and σ') and examining how often the simulations yielded μ' values greater than $10°$. The frequency of such occurrences would approximate the probability that the actual μ of a nunchuck's bend angle distribution exceeded $10°$.

Specifically, for each nunchuck image sequence, we performed the following steps 500

69
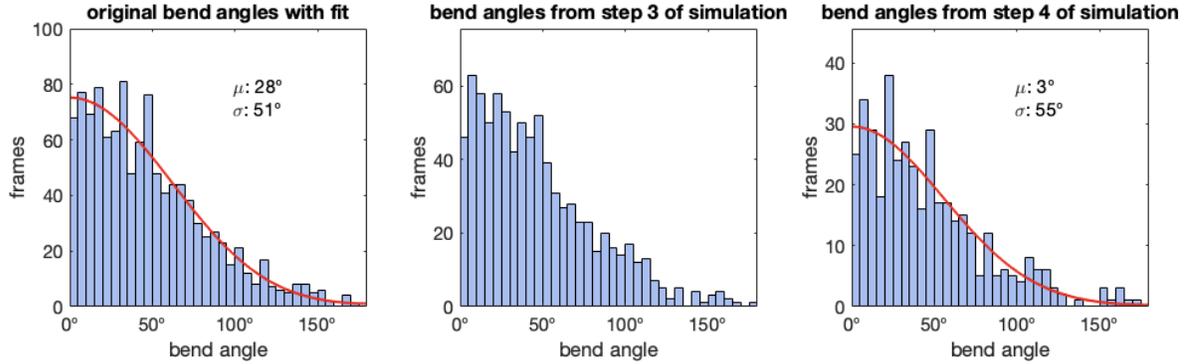
Figure 5.8: Histograms of bend angles from a single nunchuck simulation (1 of 500 Runs). From left to right: bend angles measured from the image sequence of the nunchuck, with its best TFnormal fit shown in red and fitting parameters printed; (absolute values of) simulated bend angles produced from step 3 based on the best TFnormal fit to original bend angles; bend angles produced from step 4 based on outputs of a neural network (middle), where its best TFnormal fit is shown in red and fitting parameters are printed.

times:

1. The real nunchuck's bend angle histogram was fit to a TFnormal distribution (Figure 5.8 left).

2. We have determined in previous sections that the time it takes for nunchuck bend angles to become decorrelated is ∼1 sec. Accordingly, we calculated an effective frame number for this image sequence, *e.g.* if a nunchuck was imaged at 2 Hz, then its total frame number/2 = its effective frame number, n, which represents the number of independent bend angle measurements from each nunchuck.

3. We randomly drew n bend angles from a TFnormal probability distribution, of which μ and σ were determined by the best fit in step 1. This step simulates the process of sampling a nunchuck for bend angles when the nunchuck has as true distribution of TFnormal(μ, σ). An example is shown in Figure 5.8 middle.

4. Next, we simulated the process of "measuring the bend angle from each frame of this
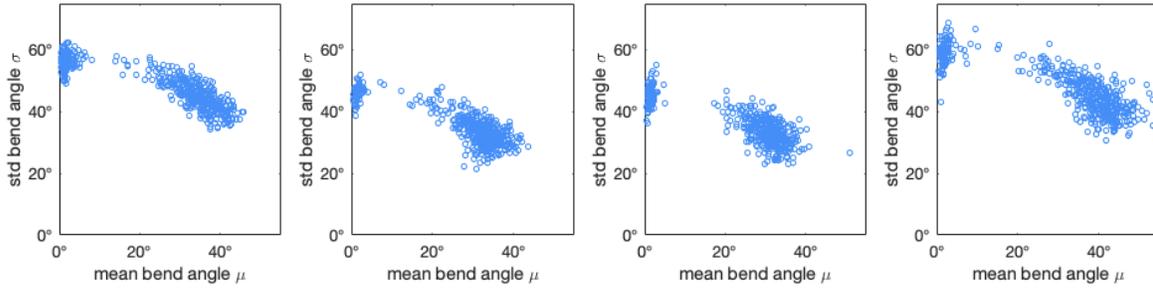
Figure 5.9: Scatter plots of σ vs. μ from 500 simulation iterations per nunchuck, for four different nunchucks. The leftmost figure corresponds to the nunchuck in Figure 5.8.

simulated image sequence by a neural network". Based on which of our our neural networks was used to analyze the original image sequence, we consulted the test result of this neural network measuring fake images generated by its own training image generation algorithm: specifically, for each simulated bend angle produced by step 3, we randomly drew a neural network test case which had the true angle from the same 5-degree bin as this simulated bend angle, and recorded the output of the neural network on this true angle. Thus, we obtained a bend angle histogram that represents the neural network analysis result of this simulated image sequence. An example is shown in Figure 5.8 right.

5. We took the absolute value of the histogram, fit it to a TFnormal distribution, and recorded μ and σ from the best fit.

This method simulates the inaccuracies and biases that emerge when sampling a finite set of data points from an underlying distribution of bend angles and then employing a neural network for the retrieval of these angles. It also takes into account the uncertainties involved in fitting a bend angle histogram to a TFnormal distribution. For each nunchuck image sequence, we have thus obtained a "probability distribution" of its true μ and σ, represented by 500 μ - σ pairs. We can then examine this probability distribution to

determine, say, how likely it is that a 500-frame nunchuck image sequence analyzed by NN3 with a best fit of $\mu = 20°$, $\sigma = 45°$ was in fact straight on average ($\mu < 10°$). Figure 5.9 displays, as examples, the outcomes of 500 simulations for four distinct nunchucks. A common observation across all four scenarios is the concentration of results along narrow, arc-like formations. These arcs are reminiscent of those observed in Figure 4.4. They correspond to regions of ambiguity in the TFnormal distribution. These arc formations encapsulate the minor variations in histogram shapes that arise from two key processes: the sampling of nunchuck bend angles (as simulated in Step 3) and the subsequent analysis of these angles using a neural network (as simulated in Step 4).

For each nunchuck then, we could examine the distribution of mean (μ) and standard deviation (σ) values from 500 simulations to establish the confidence intervals for their original TFnormal fitting parameters. For instance, consider a specific nunchuck: if over 95% of simulations yield a μ greater than 10°, we can assert with 95% confidence that this nunchuck's mean bend angle exceeds 10°. This nunchuck would be represented as a red filled triangle in Figure 5.7, with the triangle's position indicating the best-fit values of μ and σ. On the other hand, if between 68% and 95% of simulations result in a μ greater than 10°, the nunchuck is depicted as a blue filled marker in the figure. All other nunchucks are displayed as hollow blue circles. In sum, this method allows us to verify that among the 71 nunchucks in Figure 5.7, the number of nunchuck possessing a non-zero mean bend angle was far greater in proportion than would be expected by chance given the confidence levels. We conclude that the variability in the bend angle distributions of nominally identical nunchucks cannot be attributed to a TFnormal fitting artifact.

5.2   Variability May Be Intrinsic to dsDNA

The persistence of nunchuck-to-nunchuck variability in bend angle distributions, after ruling out fitting artifacts, environmental influences, and design modifications, points toward an intrinsic quality of the dsDNA linkers. It's a well-documented phenomenon in single-molecule research that molecules, ostensibly identical, can exhibit unique behaviors. This individuality has often been traced back to the presence of long-lived kinetic intermediate states, which have been shown to significantly affect the behavior of various protein molecules, leading to observable individual characteristics within molecular populations[55, 56, 57, 58]. Drawing a parallel from protein studies to our dsDNA linkers, it's plausible to hypothesize that the "kinetic intermediate state" in our context could be represented by a bent state of the dsDNA. The variability observed might then be the result of the linker transitioning into rare, meta-stable bent states. These states might not be equally sampled within our typical observation windows, which are usually less than 20 minutes.

To probe deeper into this hypothesis, we extended our observations of several nunchucks to over an hour each. Figure 5.10 shows how the bend angle distribution varies with time for four nunchucks each with an identical 37 bp linker. By comparing the bend angle distributions from distinct, non-overlapping 500-second intervals, we noticed that each nunchuck exhibited at least one interval with a bend angle distribution markedly different from others. Notably, some intervals were characterized by an unambiguously bent configuration, where the mean bend angle (μ) exceeded 40° with a 95% confidence level. These findings are not isolated instances but part of a broader pattern where many nunchucks displayed time-dependent variability in their bend angle distributions, with some maintaining non-zero mean angles for extended durations.

These extended observations reveal that the dsDNA linkers in nunchucks are dynamic
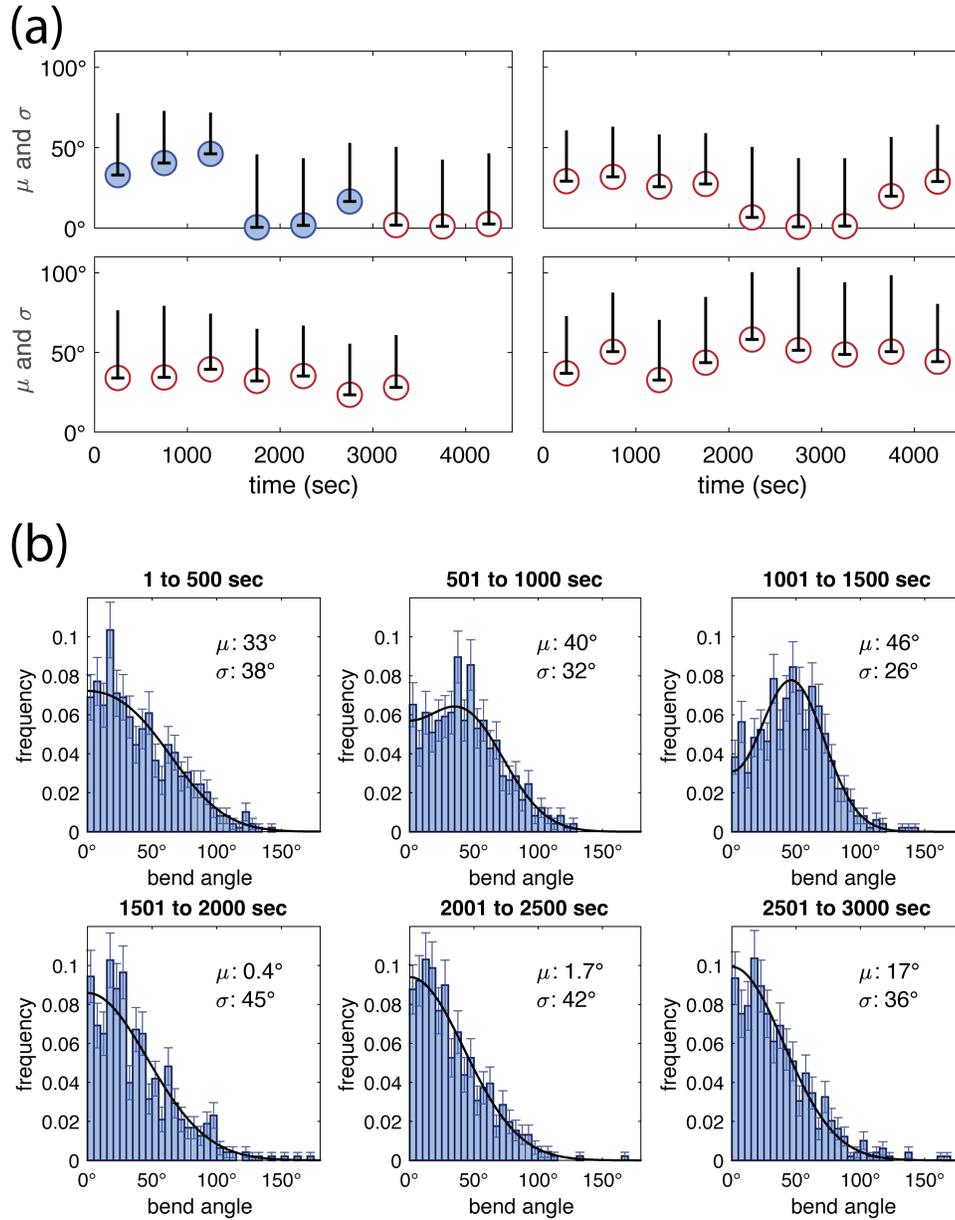
Figure 5.10: (a) The mean (μ, represented by circular markers) and standard deviation (σ, represented by lengths of black, vertical bars) of a TFnormal fit to the bend angle distributions from each non-overlapping, 500-second chunk of four 37 bp nunchucks that were observed for over one hour each. (b) Bend angle histograms from the first six 500-second chunks (blue, solid markers) of the top-left subplot in (a). The best-fitting TFnormal curves are plotted, along with their μ and σ values.

74

structures, which adopt various conformations or states over time that contribute to the observed variability in bending behavior. In the next chapter, we seek to identify and characterize the ground state and bent states of a 37 bp linker using nunchuck data.

Chapter 6

Nunchuck Bent States

6.1 The Ground State of 37 bp Linkers

If we interpret the variability in bend angle distributions as arising from a B-form structural baseline (ground state) and several meta-stable bent states, our data might reveal insights into the characteristics of these states. The ground state, being the most frequently observed condition, has a bend angle distribution that is easier to define. Specifically, it is reasonable to assume that nunchucks whose bend angle distributions are well-described by a TFnormal model with $\mu < 10°$ at a 68% confidence level (filled blue circles in Figure 5.7) are likely to have spent minimal time in bent states during observation. By aggregating the uncorrelated bend angles of these nunchucks, achieved through downsampling their time series of bend angles, and applying a TFnormal fit to the resulting combined histogram, we can discern the ground state characteristics of our 37 bp dsDNA linker. This analysis reveals a standard deviation (σ) of $46 \pm 1°$ and a mean (μ) that is close to zero, as shown in Figure 6.1(a). This observation aligns with our earlier results ($\sigma = 49 \pm 4°$) from previous chapters for 37 bp linkers, obtained by excluding "outliers" from the analysis. The exclusion had been logical when considering

that these outliers likely represent nunchucks that spent a substantial amount of time in a bent state, a condition not accounted for by the WLC model. Figure 6.1(b), in contrast, shows a similar analysis for an alternate 37 bp linker, used as a control for the 37 bp linker we have discussed so far. While the default 37 bp linker was derived from the specific sequence that binds the Integration Host Factor protein (IHF) [59], intended for potential use of nunchucks for studying protein-induced DNA bending, the alternate linker sequence has the IHF recognition site scrambled so as to act as a nonspecific binding control for forthcoming IHF research. Specifically, the 37 bp default linker sequence (GGA-GACACGGATAGGCCAATATAGCATGTGATCGGTG) was designed by removing 21 bases from the 58 bp linker (GGAGACACGGATAGGCCAATATAGCATTGCTTAT-CAATTTGTTGCAACGTGATCGGTG), which in turn was dervied from the IHF H' sequence (GGCCAAAAAAGCATTGCTTATCAATTTGTTGCACC). The 58 bp linker shares the subsequence "GCATTGCTTATCAATTTGTTG" with H', which includes the consensus sequence WATCARNNNTTR (W is A or T; R is A or G) that enables recognition by IHF, but the A_6-tract that was present in the original H' has been removed from the 58 bp linker sequence, so that the latter did not possess an intrinsic bend. The sequences of the 37 bp alternate linker, as well as that of the default one, are shown in Figure 6.2.

6.2   The Predominant Bent State of 37 bp Linkers

Having thus characterized the distribution parameters of the ground state of 37 bp linkers, we are now able to shift our focus to the bent states of these linkers. Extracting the parameters for these bent states presents a challenge. Directly fitting the bend angle histograms to a sum of multiple truncated folded Gaussian distributions isn't feasible due to the limited data points (only 36 from 5°-binned histograms spanning from 0° to
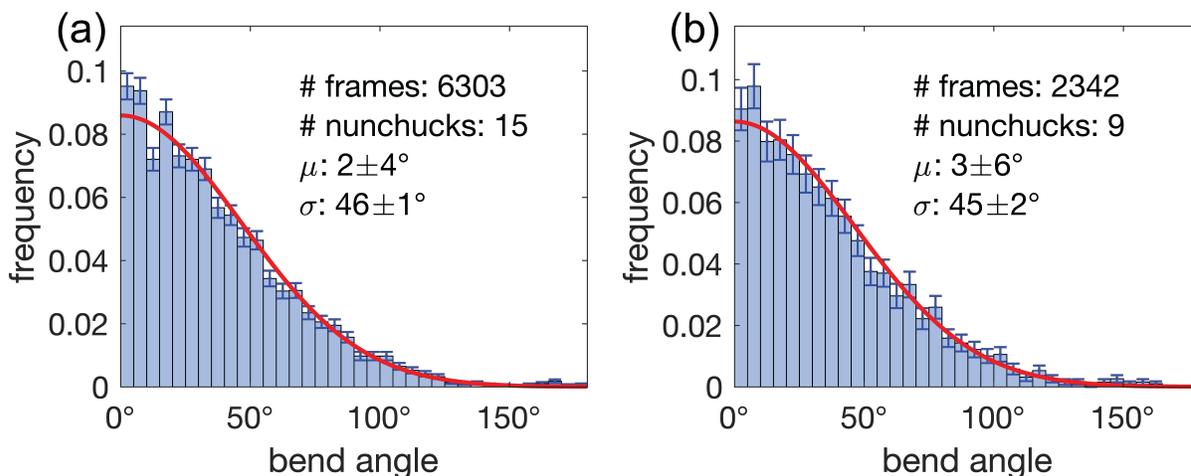
Figure 6.1:   Histograms with folded Gaussian fits (red curves) of aggregated (decor-related) bend angles from (a) 37 bp (default) and (b) 37 bp (alternate sequence) nunchucks which were well described by a folded Gaussian and had $\mu < 10°$ at 68% confidence level. Uncertainties in the mean, μ, and standard deviation, σ, were determined by bootstrapping for 500 iterations at the original sample size.

37bp_0At Linker 1	ATGCTATATTGGCCTATCCGTGTCTCCTTTGGGCAGGGCTCCGTGAATCTGGTCTGCT CTTTTGGAGACACGG
37bp_0At Linker 2	ATAGGCCAATATAGCATGTGATCGGTGTTTGGGCAGGGCTCCGTGAATCTGGTCTGCT CTTTTCACCGATCAC
37bp_alternate Linker 1	CATACGATATCGAAGCATATATCTCCATTTGGGCAGGGCTCCGTGAATCTGGTCTGCT CTTTTTGGAGATATA
37bp_alternate Linker 2	TGCTTCGATATCGTATGACTATAGCCATTTGGGCAGGGCTCCGTGAATCTGGTCTGCT CTTTTTGGCTATAGT

Figure 6.2:   Linker strand sequences for the 37 bp default (0At) and control (alternate) linkers, respectively. The sequences are color-coded following the convention of Supplementary Note S4 of our previous work[40].

180°). Attempting to fit more than two truncated folded Gaussians would result in an impractical number of free parameters (roughly 10).

Nevertheless, a viable approach emerges if we consider that a nunchuck might oscillate between only the ground state and a single bent state during the imaging process, if the observation period is kept short. Under this assumption, extracting the parameters for one bent state becomes feasible by fitting the data to a sum of two TFnormal distributions – one representing the ground state and the other the bent state. To further refine this method, we assume that the ground state's duration is at least as long as the bent states, which is reasonable as it is by definition the state with the lowest energy. This strategic limitation enhances the accuracy of our parameter extraction for the bent state, simplifying our analytical model while maintaining its effectiveness.

To achieve this, we recorded 91 short image sequences of nunchucks with the 37 bp (default) linker, each lasting approximately 5 minutes, at a sampling rate of about 10 Hertz, which is the highest rate achievable by our computer. This approach ensured that our histograms included thousands of bend angles for analysis, enough to establish a fit with high certainty. For each short image sequence, we fit its bend angle histogram to a sum of two TFnormal distributions. In this model, we assigned amplitudes C_1 and C_2 to represent the ground state and the primary bent state, respectively. The fit was structured with certain constraints: the sum of C_1 and C_2 was maintained at unity $C_1 + C_2 = 1$, and C_1 was always greater than C_2 ($C_1 > C_2$). Additionally, we anchored the ground state parameters, mean (μ_1) and standard deviation (σ_1), to the values we previously determined from Figure 6.1(a).

To verify the robustness of our fitting approach and to determine the uncertainty in best fit parameters, we carried out 1,000 bootstrapped fitting iterations for each short image sequence. In each iteration, we created a new sample by randomly selecting bend angles from the original dataset with replacement, ensuring the sample size remained
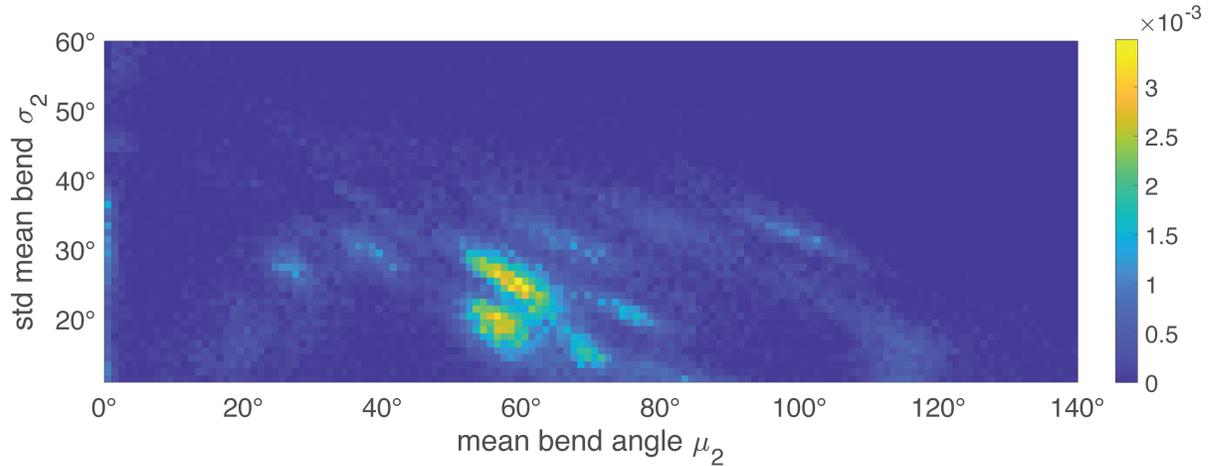
Figure 6.3:   Two-dimensional heat map histograms illustrate the frequency distribution of the mean bend angle (μ_2) and standard deviation (σ_2), parameters describing the bent state, derived from 1,000 bootstrapped iterations of 2-TFnormal fits, applied to individual nunchucks equipped with a 37 bp linker. These iterations were then selected based on the quality of the fit, with inclusion criteria set for those achieving a coefficient of determination (R^2) within 0.05 of the original bend angles' fit. The color gradient represents the density of occurrences within each bin, with warmer colors corresponding to higher frequencies, visualizing the predominant bend angle characteristics of the examined nunchucks.

unchanged. The fits for these samples started from randomly chosen values for the mean (μ_2) and standard deviation (σ_2) of the predominant bent state. This iterative method allowed us to thoroughly evaluate the stability and reliability of our fit, thereby providing a more accurate representation of the bent state's parameters.

The parameters (μ_2, σ_2) obtained from the bootstrapped fits of the 91 short image sequences — each corresponding to a different nunchuck — are collectively presented as a two-dimensional histogram in Figure 6.3. The heat map revealed a bifurcation in the behavior of the nunchucks: approximately half exhibited a widespread distribution of μ_2 and σ_2 values, suggesting the absence of a single predominant bent state throughout the brief imaging window. In contrast, the remaining nunchucks demonstrated more concentrated regions in the histogram. Notably, about one-third of the nunchucks yielded
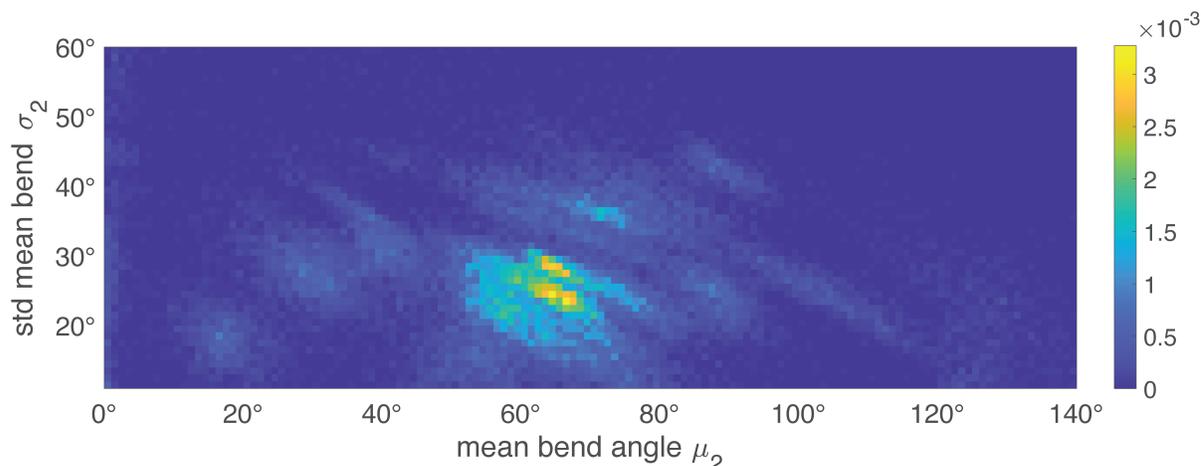
Figure 6.4: Two-dimensional heat map histograms illustrate the frequency distribution of the mean bend angle (μ_2) and standard deviation (σ_2), parameters describing the bent state, derived from 1,000 bootstrapped iterations of 2-TFnormal fits, applied to individual nunchucks equipped with a 37 bp alternate linker, acting as a control for the default 37 bp linker. Inclusion criteria similar to those of the default 37 bp linker applies. The color gradient represents the density of occurrences within each bin.

a pronounced cluster within the range of $55° < \mu_2 < 65°$ and $15° < \sigma_2 < 30°$, indicating a consistent bent state during observation. All other visible clusters within the heat map came from only one nunchuck each.

To investigate if the state characterized by $(55° < \mu_2 < 65°, 15° < \sigma_2 < 30°)$ is specific to the 37 bp linker's sequence, we repeated the study for nunchucks made with the 37 bp alternate linker, which was introduced in previous sections. Accordingly, we subjected these alternate nunchucks to the identical bootstrapping method, calibrating the analysis with their unique ground state parameters (σ_1 for the alternate linker being $45 \pm 2°$), which proved to be statistically similar to those of the default linker. The bootstrapping results for all 78 nunchucks with the alternative linker also predominantly clustered within the $(55° < \mu_2 < 65°, 15° < \sigma_2 < 30°)$ range, as depicted in Figure 6.4.

For the sake of caution, it is wise to pause and consider whether the two sequences we've selected for our nunchucks — the default 37 bp and the alternate — are adequately

representative of generic double-stranded DNA in terms of flexibility and propensity for kinking. If the sequences were to exhibit unusual flexibility or kink propensity, the predominant bent state we observed might not be applicable to other dsDNA sequences. To address this, we utilized the model developed by Basu *et al.*[60], which was developed based on high-throughput DNA cyclizability data measured via the loop-seq method. Their model was able to accurately predict DNA bendability based on local sequence and epigenetic modifications. By implementing their MATLAB program, we assessed the cyclizability of our chosen sequences. The results indicated that, compared to average dsDNA, our sequences exhibit standard flexibility, neither unusually high nor low. This finding alleviates concerns that we had selected particularly anomalous DNA sequences for our study. Furthermore, the study by Park *et al.* [61] delves into the molecular dynamics and theoretical modeling of DNA kinking. They specific finding lies in the significant increase in kink propensity in sequences featuring TA dinucleotide repeats flanked by GC steps, a characteristic present in our alternate linker sequence, but not in the default linkers. Despite this, the predominant bent state cluster was observed in both the default and alternate linkers with no significant difference, suggesting that this bent state is not merely an artifact of sequence-induced kinking. By considering these two pieces of research, we can confidently assert that the default and alternate 37 bp linker sequences used in our study are not outliers in terms of bendability. The consistency of the predominant bent state across these two distinct sequences is indeed a common feature among dsDNA of a similar length.

While Figure 6.3 indicates the existence of a predominant bent state around ($55° < \mu_2 < 65°, 15° < \sigma_2 < 30°$) as conspicuous clusters in the heatmap, it stops short of providing numerical values for μ_2 and σ_2. Recognizing that not all nunchucks manifests this feature in the heat map and, therefore, may not all be useful for extracting the attributes of this bent state, we excluded from subsequent analysis the nunchucks for
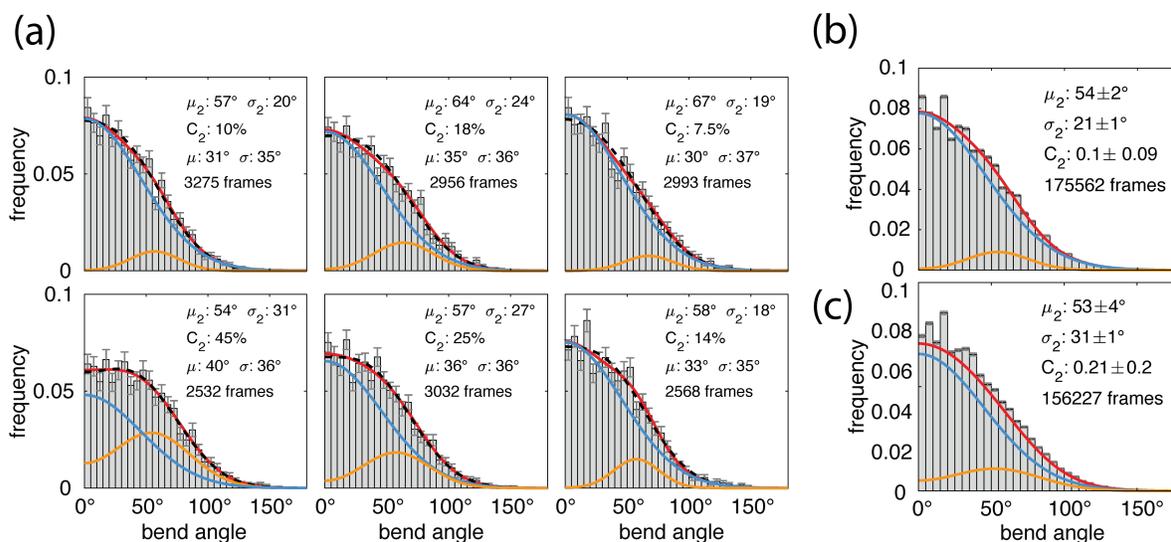
Figure 6.5:   (a) Presented here are the bend angle histograms for six 37-bp nunchucks which produced a prominent, bright cluster in its 2D histogram of μ_2 vs. σ_2. For each subplot, the black, dashed curve shows the single TFnormal fit, with parameters (μ and σ) displayed. The 2-TFnormal fit is illustrated by the red solid curve, which is composed of the ground state (blue curve) and a bent state (orange curve). The ground state parameters (μ_1 and σ_1) are fixed to the values determined from Figure 6.1a. The parameters characterizing the bent state (μ_2, σ_2, and the amplitude C_2), as well as the total number of data points, are annotated on the histograms.   (b-c) Aggregate histograms of bend angle data are shown for all selected nunchucks with 37 bp default linker (b) and 37 bp alternate linker (c), specifically excluding those nunchucks whose predominant bent state did not fall within the specified parameter range ($55° < \mu_2 < 65°, 15° < \sigma_2 < 30°$). The combined 2-TFnormal fit (red curve) in each histogram represents the sum of the ground state (blue curve) and a bent state (orange curve). The bent state parameters (μ_2, σ_2, and the amplitude C_2) are provided in the figures, with the number of data points included. Uncertainties were calculated using a bootstrapping method to ensure robust statistical interpretation.

which the main bright cluster was away from the range of $(55° < \mu_2 < 65°, 15° < \sigma_2 < 30°)$ - differing by at least $10°$ in all directions, suggesting a different bent state dominated during their observation. Hence, 23 out of 91 nunchucks with the original 37 bp linker were excluded from the aggregate. These remaining nunchucks are expected to be well characterized by the sum of two TFnormals, where one represents the ground state and the other corresponds to the predominant bent state. Figure 6.5 showcases six such nunchucks: the red solid curve illustrates the composite 2-TFnormal fit, which is equal to the sum of the ground state (blue curve) and the bent state (orange curve). On the other hand, the black dashed line represents the best single TFnormal fit to the bend angle histogram, and the best fit parameters always had $\mu > 10°$ (often $\mu \approx 30°, \sigma \approx 35°$). Notably, the black (TFnormal) and red (2-TFnormal) fits appear visually similar, both aligning closely with the histogram data. This resemblance is significant yet baffling, as it is the very reason why this predominant bent state has remained elusive until now, due to the satisfactory fit provided by a single TFnormal model. Nonetheless, we know that the single TFnormal model is inappropriate since it conflicts with the characteristics of the ground state, which has a mean bend angle $\mu = 0°$, whereas the TFnormal1 fit always yields $\mu > 10°$. This crucial distinction will be further discussed in the following section. For now, we note that the large variability among individual nunchuck distributions observed in Figure 5.7 was precisely a result of fitting the bend angle distributions to one single TFnormal, while in reality the distributions should have been described by a sum of two states. The two TFnormals, representing the ground state and the predominant bent state, can collectively produce a histogram that fits well to one single TFnormal with its best fit parameters μ being either close to $0°$ (if the ground state dominates) or $60°$ (if the bent state dominates), or landing somewhere in the $20° < \mu < 40°$ range (if both states are present), which gives the impression in Figure 5.7 that nunchucks display a continuous spectrum of bending behaviors.

For a definitive extraction of μ_2 and σ_2, we aggregated the histograms of the 68 selected nunchucks, resulting in a collective histogram comprising 175,562 frames of decorrelated images, as shown in Figure 6.5(b). Fitting this histogram with a 2-TFnormal model while fixing μ_1 and σ_1 to reflect the ground state yielded $\mu_2 = 54 \pm 2°$ and $\sigma_2 = 21 \pm 1°$, which characterize the 37 bp default linker's predominant bent state. Uncertainties were obtained from 500 iterations of bootstrapping on the aggregated dataset. The fitting curves follow suit, with the orange curve representing the bent state and the blue curve the ground state, culminating in the red curve, which is the 2-TFnormal fit to the aggregated data. These results further indicate that the bent state accounts for 10% of the aggregated angles (*i.e.*, $C_2 = 0.1$). We conclude that the overall fraction of time spent in this bent state is $0.1 \times 68/91 = 7.5\%$ for this particular linker. Similar analyses was performed for the 37 bp alternate linker, and the results as shown in Figure 6.5(c) are $\mu_2 = 53 \pm 4°$ and $\sigma_2 = 31 \pm 1°$. While the two linkers differ in the standard deviations of bend angles, the mean bend angles are in excellent agreement with each other, corroborating the idea that this bend state is independent of linker sequence.

The bent state near $(55° < \mu_2 < 65°, 15° < \sigma_2 < 30°)$ is the only one adequately represented in our data to permit a thorough analysis using our 2-TFnormal model. To extend this level of characterization to other bent states, which can be seen in Figure 6.3 as dimmer clusters or blobs, the modest frequency of their occurrence in our dataset implies the necessity for a larger sample size. To gather a statistically significant representation that could verify and detail additional bent states, an expanded number of nunchucks, substantially more than the 91 currently examined, would be required.

6.3   Obscurity of the Predominant Bent State

So far, based on the assumption that nunchucks imaged for a short period primarily explore a ground state plus a predominant bent state, we were able to identify this predominant bent state by fitting individual nunchuck bend angles to a sum of two TFnormal distributions, provided that one of them represents the ground state and is constrained to have a greater amplitude ($C_1 > C_2$, where C_1 and C_2 are amplitudes of the two Gaussians). From the constrained fits, results were found to aggregate around $\mu_2 \sim 60°$ and $\sigma_2 \sim 25°$, revealing a predominant bent state. However, here we show in Figure 6.6 that when the fit was performed without the constraint $C_1 > C_2$, the best fit for most nunchucks clustered around $\mu_2 \sim 30°$ and $\sigma_2 \sim 30°$. Notably, all fits in this cluster have $C_2 \gg C_1$, i.e. the ground state has near-zero amplitude, which contradicts our assumption that the ground state is the primary bend state. Hence, clusters at $\mu_2 \sim 30°$ and $\sigma_2 \sim 30°$ are artifacts of the TFnormal distributions, and do not indicate a bent state; rather, they arise simply because for nunchucks exploring a ground state plus a predominant bent state at $\mu_2 \sim 60°$ and $\sigma_2 \sim 25°$, the bend angle histograms may be well fit to one single TFnormal at around $\mu_2 \sim 30°$ and $\sigma_2 \sim 30°$.

We use simulations to verify that the ground state of 37 bp nunchucks, alongside a bent state approximating $\mu_2 \sim 60°$ and $\sigma_2 \sim 25°$ with smaller amplitudes, could collectively resemble a TFnormal distribution with parameters $\mu_2 \sim 30°$ and $\sigma_2 \sim 30°$. The simulation methodology, akin to that described in earlier sections to determine the confidence in TFnormal best fit parameters of 37 bp individual nunchucks, imitates the process of sampling a finite number of images from a nunchuck with an underlying bend angle distribution, analyzing those images with a neural network, and fitting the measured bend angles to a distribution. Unlike in the earlier simulation case, the underlying distribution here is the sum of the ground state and a bent state with $\mu_2 = 60°$ and
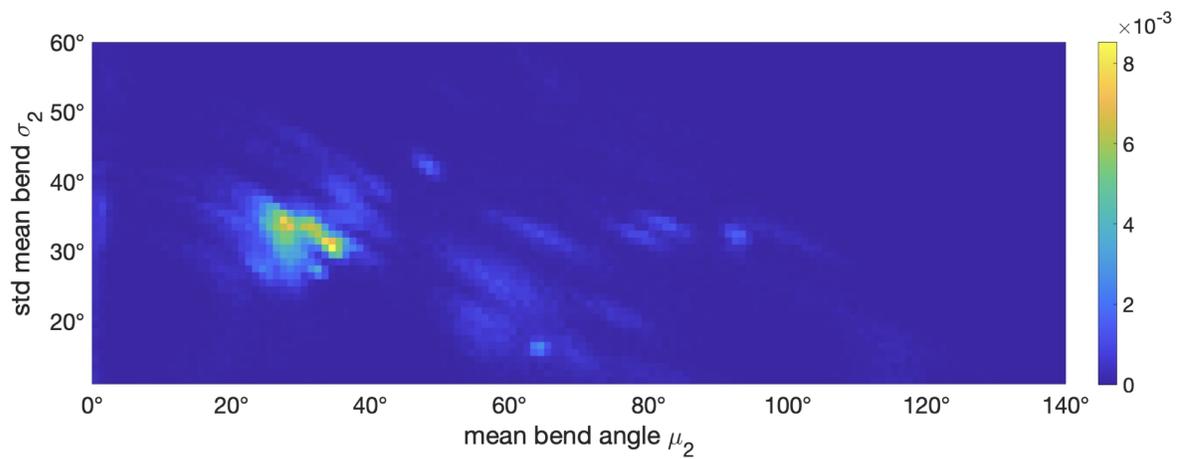
Figure 6.6:   A two-dimensional histogram as a heat map, illustrating the frequency distribution of μ_2 and σ_2 from bootstrapped 2-TFnormal fits applied to nunchuck data with the 37 bp default linker. The parameters C_1 and C_2 were left unconstrained during the fitting process. A prominent concentration of fit outcomes clusters around $\mu_2 \sim 30°$ and $\sigma_2 \sim 30°$, which seems to suggest a bent state located around these parameters. Yet, the amplitude of the ground state (or contribution ratio), C_1, is found to be negligible across all nunchuck samples, which contradicts the expected predominance of the ground state over the bent state in frequency.
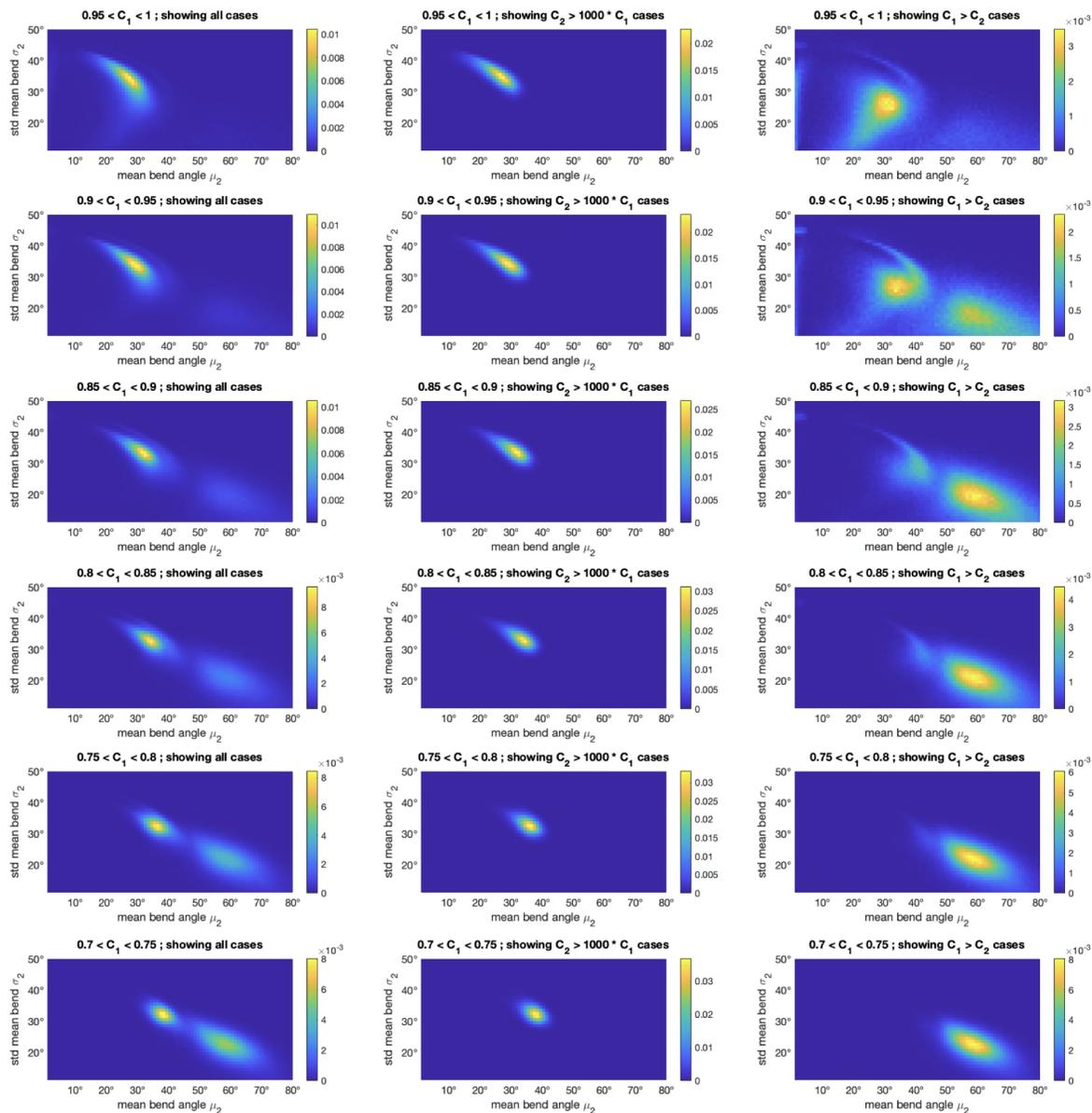
Figure 6.7: The two-dimensional histograms presented here depict the distribution of μ_2 and σ_2 from 2-TFnormal fitting of simulated data for nunchucks with 37 bp default linkers. These simulations reflect the exploration of the ground state and the predominant bent state. The histograms are arranged with the top to bottom axis representing results from simulations with varying ranges of C_1, and the left to right axis categorizing the results based on the comparative magnitudes of C_2 to C_1: encompassing all fits, solely fits where $C_2 \gg C_1$, and exclusively fits where $C_2 > C_1$. This organization allows for an intuitive comparison of how the contribution ratios affect the observed distributions of the bend state parameters.

$\sigma_2 = 25°$, and the histogram of bend angles measured by the neural network is fit to the sum of two TFnormals, one of which is set to the distribution of the ground state and is constrained to have a greater amplitude.

The details of the simulation are as follows:

1. We constructed a probability distribution by summing two TFnormals, as described by the equation below. We set μ_1 and σ_1 to ground state values of 37 bp default linker nunchucks; $\mu_2 = 60°$ and $\sigma_2 = 25°$ (representing the predominant bent state). The amplitudes of the two TFnormals are C_1 and C_2 respectively, and for a normalized distribution $C_1 + C_2 = 1$. Since the bent state is expected to be less populated than the ground state, we ran simulations with C_1 between 0.6 and 1 so that $C_1 > C_2$ for all cases.

$$
\begin{aligned}
&\frac{C_1}{\sqrt{2\pi\sigma_1^2}} \left(e^{-\frac{(-\theta-\mu_1)^2}{2\sigma_1^2}} + e^{-\frac{(2\pi-\theta-\mu_1)^2}{2\sigma_1^2}} + e^{-\frac{(\theta-\mu_1)^2}{2\sigma_1^2}} + e^{-\frac{(-2\pi+\theta-\mu_1)^2}{2\sigma_1^2}} \right) \\
&+ \frac{C_2}{\sqrt{2\pi\sigma_2^2}} \left(e^{-\frac{(-\theta-\mu_2)^2}{2\sigma_2^2}} + e^{-\frac{(2\pi-\theta-\mu_2)^2}{2\sigma_2^2}} + e^{-\frac{(\theta-\mu_2)^2}{2\sigma_2^2}} + e^{-\frac{(-2\pi+\theta-\mu_2)^2}{2\sigma_2^2}} \right)
\end{aligned}
\tag{6.1}
$$

2. We randomly drew n bend angles from this probability distribution. This step simulates the process of sampling a nunchuck for bend angles when the nunchuck has a true distribution as given by the equation above.

3. Next, we simulated the process of "measuring the bend angle from each frame of this simulated image sequence using a neural network". We randomly picked one of our four neural networks, NN1 through NN4, and consulted the test result of this neural network measuring fake images generated by its own training image generation algorithm (described in earlier chapters): specifically, for each simulated bend angle produced by step 3, we randomly drew a neural network test case which

had the true angle from the same 5-degree bin as this simulated bend angle, and recorded the output of the neural network on this true angle. Thus, we obtained a bend angle histogram that represents the neural network analysis result of this simulated image sequence.

4. We took the absolute value of the histogram, fit it to a sum of two TFnormal distributions as described in the equation above, with μ_1 and σ_1 values constrained to be those of the ground state, and recorded μ_2 and σ_2 from the best fit. We did not constrain fitting parameters C_1 and C_2 in any way.

5. We ran 500 such simulations for each range of C_1 of the underlying distribution: [0.6, 0.7], [0.7, 0.75], [0.75, 0.8], [0.85, 0.9], [0.9, 0.95] and [0.95, 1].

Results of the simulations for various ranges of C1 are plotted in Figure 6.7 subplots. We show distributions of μ_2 and σ_2 from either all fits, only fits with $C_2 \gg C_1$, and only fits with $C_1 > C_2$. For cases with $C_2 \gg C_1$ (middle column), the ground state had near-zero amplitude, $i.e.$ the result could be well-described by one single TFnormal with μ between 30° and 40°, consistent with our interpretation of Figure 6.6 that the cluster located at $\mu_2 \sim 30°$ and $\sigma_2 \sim 30°$ was due to a bent state at $\mu_2 \sim 60°$ and $\sigma_2 \sim 25°$. As C_2 increases (from top to bottom), fits were increasingly able to correctly recover the bent state at $\mu_2 \sim 60°$ and $\sigma_2 \sim 25°$ provided that $C_1 > C_2$ (right column).

6.4   Interpreting the Predominant Bent State

Among the array of nunchucks imaged, we have identified a notable subset with the 37 bp default linker that consistently exhibits a bent state with mean bend angle at around 60°, seemingly lacking the ground state entirely. In essence, these nunchucks appeared to be persistently bent during our entire observation period. Figure 6.8 showcases

examples from this intriguing group. It is conceivable that these particular nunchucks have somehow become stabilized in this bent state that the others only sampled transiently, or they may be the expected subset of events in the tail of the exponentially distributed lifetimes of a bent state with a characteristic residence time of several minutes. Nonetheless, this latter hypothesis seems improbable. If such longevity for the bent state were typical, we would expect to see some nunchucks displaying dual peaks in their bend angle distributions, one at $0°$ and another at $60°$, signifying substantial populations in both states. Yet, our observations have never encountered this. All nunchucks we have studied have shown a much smaller amplitude for the bent state relative to the ground state.

Furthermore, the distribution of the best fit parameters μ and σ for these nunchucks spreads widely across the predominant cluster Figure 6.3. This spread implies that the state characterized by ($55° < \mu_2 < 65°, 15° < \sigma_2 < 30°$) does not correspond to a singular bent conformation but may instead be a composite of several bent states. These states, while closely aligned in angle, are distinguished by considerable energy barriers, preventing easy transitions between them and contributing to the complex landscape of nunchuck bending behaviors.

The observed mean bend angle of the predominant bent state in our nunchucks intriguingly aligns with the local minimum of the DNA bending potential outlined by Du *et al.*, centered between $50°$ and $60°$, a detail they derived by aligning their analyses with experimental data from protein-DNA complex structures[33]. This local minimum, as explained by Drozdetski *et al.*, is responsible to the energy-minimizing influence of kinked DNA states[62], suggesting a potential parallel to the bent states we encounter in our nunchuck linkers. The presence of a $54°$ bent state suggests a localized kink between adjacent base pairs, yet it does not inherently explain the observed increased rigidity throughout the entire linker ($\sigma_2 = 21°$ or $31°$ as opposed to $46°$). Additional research
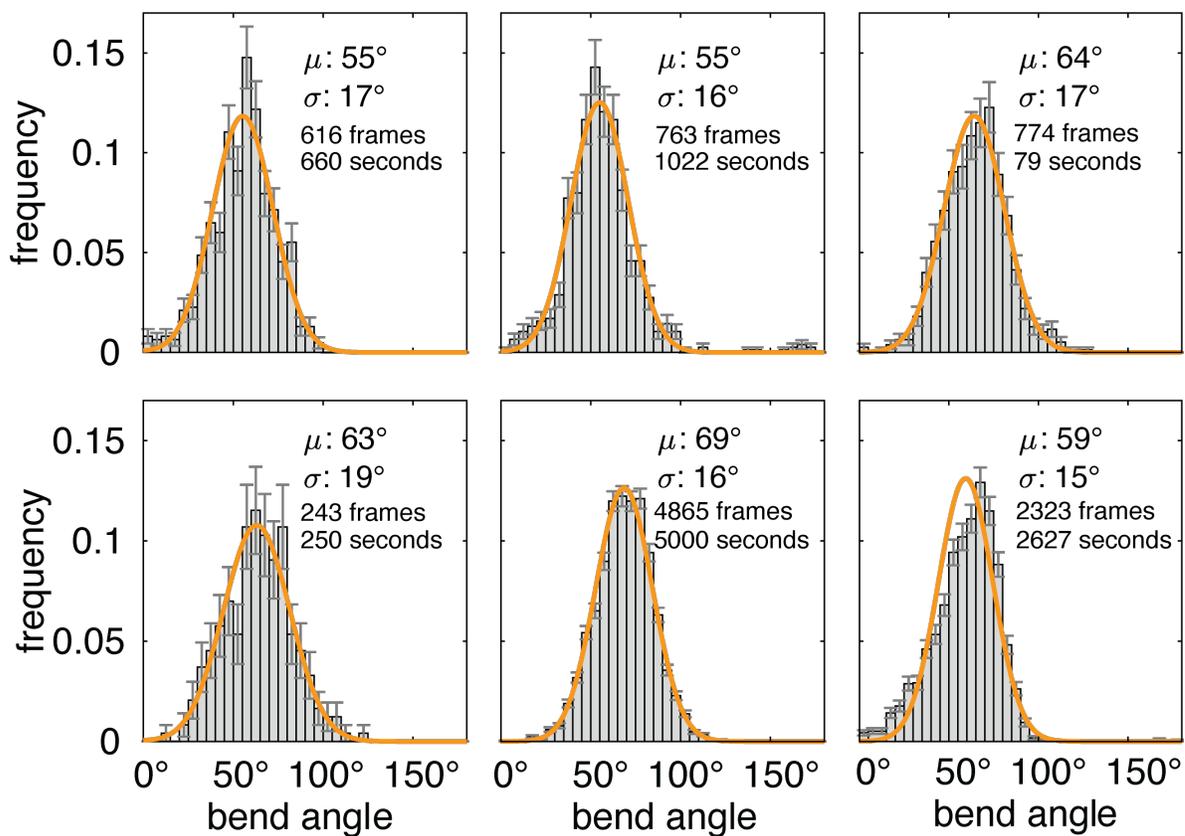
Figure 6.8: Bend angle histograms for six distinctive nunchucks, each constructed with the same 37-bp default linker and characterized by a notable absence of a ground state, instead revealing a pronounced bent state with a mean angle (μ) between 55° and 69°. Overlaying the histograms are the best TFnormal fit curves (shown in orange), with each fit's specific parameters (μ and σ) displayed in figure. The number of data points used in the histogram and the total duration of imaging are shown in each subplot.

is necessary to clarify why such a stiffened state accompanies the bend. For example, should the kink manifest uniformly at any point along the strand, we would anticipate detecting the 54° bend irrespective of the linker's length, with an increasing likelihood — mirrored by the value of C_2 — corresponding to longer linkers. In addition, further studies are required to determine the number of base pairs involved in the bend.

Our study has yielded definitive evidence for the natural occurrence of spontaneous, localized bending within short segments of double-stranded DNA, marking the first quantification of the degree and durability of states. This phenomenon, often referred to as "kinking," has long been hypothesized as a key factor underlying the anomalously high cyclization rates of short dsDNA, a topic of considerable debate within the scientific community for over two decades[63, 29, 64, 65]. The concept that DNA cyclization is augmented by kinked configurations has been increasingly supported by recent literature. For example, advanced mathematical models incorporating the allowance for kinked states have successfully replicated the observed high cyclization probabilities (known as j-factors) as well as the seemingly reduced persistence lengths that have been a source of puzzlement in experimental observations[31, 32]. Furthermore, coarse-grained simulations of DNA using oxDNA have identified kinks as a critical driver behind the enhanced cyclization rates in shorter DNA strands[66].

In a landmark study, Drozdetski *et al.* demonstrated how kinked DNA states could lower the total energy of a DNA loop, drawing on the presence of a non-convex region in the DNA's bending potential curve. This finding aligns with our observations from the nunchuck assay, where we've identified a predominant bent state characterized by a significant bend coupled with decreased flexibility. Such a state suggests the presence of a meta-stable kink which could act as a structural facilitator for DNA cyclization, providing a missing piece in the puzzle of DNA's mechanical behavior. The relatively large bend and small flexibility of the predominant bent state observed during the nunchuck

assay observations are, at least, in line with the notion of a meta-stable kink that could potentially enhance cyclizability.

Chapter 7

Prospects of the DNA Nunchuck

The DNA nunchuck has proven to be a practical instrument for single-molecule measurement of DNA bending, providing accessibility to any laboratory equipped with conventional fluorescence video microscopy. In this dissertation, we have outlined critical aspects of nunchuck design. Our findings reveal that for linkers with lengths between 27 and 58 base pairs, an "out-of-phase" configuration, positioning the arms on opposite sides of the linker, is necessary to prevent steric hindrance and ensure that linker flexibility is accurately represented. We calibrated the nunchuck using linkers that contained no intrinsically bent motifs and determined the effective linker length, facilitating future measurements and analyses of linker flexibilities. Additionally, our study confirms that nunchucks can accurately measure the static bend inherent to the linker. This was demonstrated by incorporating phased A_6-tracts into the nunchucks and successfully retrieving the expected $17°$ bend per A_6-tract from the nunchuck data. Moreover, we were able to quantify the A-tract stiffness using this data.

Our study also delved into the intriguing observation that nunchucks with identical structures display different behaviors. Despite extensive efforts to eliminate this variability, we ultimately attributed it to an intrinsic property of the linker. This assumption led

us to uncover a 54° bent state in nunchucks with 37 bp linkers, characterize its parameters, and confirm its sequence independence. Although the exact cause of this bent state remains unclear, it correlates with the enhanced cyclizability observed in short dsDNA by previous studies. This finding could potentially explain a long-standing debate in the field and add valuable insights to ongoing discussions.

7.1   Studying Intrinsic Bend of DNA Motifs

DNA motifs have become increasingly significant not only for their biological roles in gene regulation and protein recognition but also as fundamental elements in the burgeoning field of DNA technology, where they serve as building blocks for DNA nanostructures. This growing importance underscores the need to understand the mechanics and dynamics of DNA motifs, sequences, and other special constructs. The flexibility and conformation of DNA nicks, for instance, have been subjects of longstanding debate among researchers. While some studies suggest that nicked DNA retains the B-form structure characteristic of intact DNA strands [67, 68, 69], others propose that nicks disrupt this structure, leading to local kinks or bends [70, 71, 72]. Our research has contributed to this discussion by examining the impact of linker nicks. In earlier chapters, we have compared nunchuck bend angle data from nunchucks made with no-nick linkers, with ligated linkers (sealed by ligase with imperfect efficiency, leaving some nicks unligated), and with unligated linkers (each having two nicks). Our findings revealed no significant differences in flexibility, suggesting that nicks do not significantly influence the bending dynamics of linkers. Other motifs and constructs that have attracted research interest include DNA gaps [69, 73] and triplexes [74, 75], as well as more complex structures like bulges, bubbles, and Holliday junctions [76]. For example, the work of Schreck *et al.* employed oxDNA simulations to characterize the bending of DNA bulges, highlight-

ing their relevance in DNA nanotechnology [15]. These motifs can significantly affect the properties and structural behavior of DNA nanostructures, offering opportunities for fine-tuning and controlling self-assembly processes. Additionally, various studies have used diverse methodologies to investigate the bend and flexibility of bulges with different base counts [77, 78, 79, 80, 81], as well as the properties of bubbles comprising different bases and lengths [82, 83, 84, 85, 65].

The versatility of our nunchuck design allows for its effective adaptation in characterizing various DNA motifs. We have recently exemplified this adaptability by creating nunchucks with linkers that incorporate bulges and bubbles. Specifically, we developed nunchucks with 37 bp linkers that include bulges of either 3 or 6 thymine bases, as well as bubbles with either 2 or 6 thymine bases. For each of these configurations, we have gathered data from a minimum of 20 image sequences. The analysis of this data is still pending and thus will not be included in this dissertation. Nonetheless, the straightforward and efficient process of constructing these nunchucks underscores their potential utility in studying such DNA structures.

7.2   Studying Intrinsic Bend of RNA Motifs

RNA nanotechnology is rapidly advancing as a highly promising field, offering greater potential than DNA nanotechnology due to its unique properties and versatility. While DNA has been the cornerstone of nucleic acid nanoconstruction for decades, recent insights into RNA's structure have led to the burgeoning field of RNA architectonics, aimed at creating diverse nanostructures. RNA stands out with its modularity, complex tertiary folding akin to proteins, and functional versatility, evidenced in natural RNAs that facilitate critical cellular processes. Unlike DNA, which primarily utilizes secondary structures for self-assembly, RNA's potential lies in its stable tertiary structures, enabling

the design of intricate and reversible nanoarchitectures[86]. These distinctive properties of RNA, including its ability to form complex shapes and perform catalytic functions, position it as a more innovative and adaptable material in nanotechnology compared to DNA.

Recent scholarly attention has converged on the multifaceted properties of RNA motifs and their interplay with proteins, underpinning a surge of studies in this dynamic field. As examples, bulges within DNA and RNA, while commonplace, serve as critical recognition sites for specific proteins, consequently influencing the geometry and stability of RNA duplexes, with their significance expansively studied [87]. The advent of double-kink-turn RNA motifs as fundamental building blocks for diverse nanostructures illustrates RNA's structural versatility [88]. The dynamic nature of RNA four-way junctions, with their inherent flexibility and stacking variations, has been crucial in unraveling the complexities of RNA folding and tertiary structure [89, 90]. The functional and structural insights into RNA kissing loops have propelled understanding of genomic RNA dimerization in retroviruses [91]. Ingeniously designed motifs such as corner motifs and right-angle motifs have been exploited to engineer artificial RNA nanomaterials, extending the potential of RNA into the realms of nanorings, nanotubes [92], and even RNA squares [93]. A-minor junctions control coaxial stacking of helices, aiding in the folding and structural prediction of RNA, which in turn bolsters RNA nanobiotechnology and synthetic biology [94]. The adaptability of the tetra-U RNA motif has been harnessed to create finely tunable and stable nanostructures, marking a notable advance in RNA nanotechnology and its medical applications [95]. The continuous exploration of RNA motifs and RNA-protein interactions, highlighted by Shibata *et al.*'s protein-driven RNA nanostructures with therapeutic potential [96] and Saito *et al.*'s translational regulator systems [97], showcases the burgeoning interest and the critical importance of this field. These studies illuminate the extensive versatility and the promising horizon of RNA tech-

nology, underlining the necessity of rigorous inquiry into the bending and interactional behaviors of RNA.

The burgeoning field of RNA nanotechnology has prompted us to explore the measurement of RNA motif bending using the nunchuck technique. In fact, we have integrated a 27-bp RNA linker into our nunchuck construct with success, despite RNA's known vulnerability to degradation. To ensure the integrity of these RNA-incorporated nunchucks, which are known to be susceptible to enzyme degradation, we implemented rigorous laboratory protocols, including maintaining a clean work environment. However, we chose not to present the data here since we adopted a nunchuck seed design which deviates from that presented in this dissertation. Specifically, we utilized two different seed monomers. Although both seeds were derived from the same p3024 scaffold, the scaffold regions for folding in the second seed was shifted to ensure orthogonal staple, adapter, and linker attachment sites, enhancing the structural fidelity of the nunchuck assembly. This innovative seed design, while ensuring the correct assembly of nunchuck components, has not been calibrated to determine its effective linker length, and so we cannot reliably extract the RNA linker flexibility from data. Despite these challenges, the successful assembly and imaging of these RNA nunchucks underscore the viability of using RNA linkers in this assay. Moving forward, it is conceivable to substitute the RNA linker with various RNA motifs of interest, such as those previously discussed. Given the significant interest in the inherent bending angles and flexibility of these motifs from both a biological and nanoengineering perspective, our nunchuck assay provides a promising platform for further study.

7.3   Studying Protein-DNA Interactions

DNA's interaction with various proteins, which can mechanically modify it, is of great importance in biological research. Many proteins bind specifically to DNA sequences with recognizable sites, such as IHF, known for making the DNA do a "U-turn", *i.e.,* inducing a nearly 180° bend in DNA[98]. The mechanism of IHF's recognition, binding, and subsequent bending of DNA has been extensively studied[99]. Other proteins, like HMGB1, can bind to DNA with structural anomalies like bends or kinks, and can induce a mean bend angle of 67° upon binding[100]. In contrast, proteins such as RecA bind indiscriminately to DNA or RNA, "coating" the DNA backbone and enhancing its rigidity. Studies have shown that RecA can form a "nucleoprotein filament" on ssDNA or dsDNA, binding approximately one RecA monomer per three nucleotides[101].

In our lab, we are currently exploring the nunchuck technique's potential for investigating protein-DNA interactions. We have conducted preliminary trials with HMGB1 on nunchucks with linkers containing bulges or bubbles, creating defects for HMGB1 binding. Additionally, we have imaged the bending motion of nunchucks with linkers made from the IHF-specific sequences in an environment containing various concentrations of IHF. We found that sealing all nicks in the nunchuck by ligation, including those in the arm nanotubes (provided the nanotube tiles were kinated beforehand), enhances the nunchuck's robustness, maintaining structural integrity from 4°C to 55°C. This thermal stability is crucial for protein-DNA interaction studies, as proteins often require specific temperatures to avoid degradation (4°C) or to mimic physiological environments (37°C). Although our initial tests with IHF and HMGB1 haven't yet shown induced changes in bending, this is likely due to not having reached the necessary ionic and temperature conditions for optimal protein-DNA interaction. We are currently identifying these optimal conditions for IHF through Electrophoretic Mobility Shift Assay (EMSA) studies

(publication pending) and plan to apply them to the nunchuck assay soon. Our goal is to observe the 180° bend caused by IHF using the nunchuck, demonstrating its potential as a new tool for visualizing protein-DNA interactions at a single-molecule, real-time level. RecA is another promising subject for nunchuck studies, especially given its easy integration with DNA nanostructures like DNA origami, as demonstrated by Schiffels *et al.*[102]. The RecA-DNA complex, known for its smaller twist per nucleotide (20° instead of 35°), increased length (1.5 times), and greater persistence length (about 10-fold)[102], can provide another path to calibrate and validate the nunchuck for measuring protein-DNA interactions.

7.4   Limitations and Improvements

The DNA nunchuck, while a powerful tool for probing the intrinsic bends and flexibility of DNA linkers, has certain inherent limitations. The nunchuck's design confines the range of linker lengths to between 27 and 58 base pairs. The lower limit is constrained by the thermal stability required for base-pairing at the linker-seed interface, as detailed in our previous publication[40], where strategic design allows for shorter linkers through internal nicks within the seeds. The upper limit, however, is imposed by financial considerations, as longer linkers increase production costs. While it is technically feasible to construct nunchucks with longer linkers, surpassing the linker's persistence length leads to a reduction in the correlation between the orientations of the nunchuck ends. This, coupled with the higher likelihood of the nunchuck arms colliding, adds complexity to the interpretation of bend angle histograms.

An additional constraint of the nunchuck assay is its temporal resolution. With a typical observational frame rate of around 10 Hz, the nunchuck assay lags significantly behind the millisecond to nanosecond resolution achievable by other single-molecule tech-

niques. Although the frame rate is not a restrictive factor, as we could employ higher frame rates to monitor nunchucks, the bulky arms inherent to the nunchuck design inherently slow down the movement of the linker. This reduced mobility impacts the temporal resolution of the observations we can make. As a result, the nunchuck assay's primary focus has been on analyzing the distribution of bend angles rather than the dynamics of the linker, the latter of which remains an area yet to be fully understood.

Enhancements to the nunchuck design and assembly protocol could yield improvements in cost-efficiency, assembly speed, yield, imaging simplicity, and bend angle measurement accuracy. The current seed design, derived from the work of Schulman *et al.* [41], uses a large, ∼7000-base scaffold from the readily available M13mp18. Such a choice proves costly, requiring 72 distinct staples, all of which must be purified by PAGE or HPLC, and also increases the potential for errors. Our laboratory has shown (publication pending) that much shorter seeds are not only viable but also reduce complexity and expense. Future iterations should consider smaller seed constructs to economize on staple requirements. Furthermore, adopting more rigid nanotube architectures could rectify issues arising from the flexibility of the current 12-helix nanotube design. The inherent curvature of the nanotube arms may introduce inaccuracies in imaging analysis, as the bend angle of the linker is inferred from the orientation of these arms. By employing stiffer nanotube structures, we can reduce this source of error and achieve more accurate measurements of the linker's end-to-end bend angle.

In the laboratory, one of the primary challenges with the nunchuck technique is the unpredictability of the dynamic imaging process. The non-uniform thickness of the gap created by PEG-coated coverslips and glass slide surfaces complicates the 2D confinement of nunchucks. Upon introducing a new flow cell sample under the microscope, it often requires considerable effort to find a region where nunchucks are adequately confined and in sharp focus. This trial-and-error approach is time-intensive, with no guarantee

of finding an optimal viewing area in a given sample. Once an area is found, nunchucks tend to drift, necessitating continuous manual stage adjustments throughout the imaging session to keep the nunchuck within the field of view, a process that is both labor-intensive and a deterrent for new users familiarizing themselves with the technique.

Improvements to the flow cell design could streamline and simplify this process. One potential solution is the adoption of prefabricated nano-arrays with shallow pits or welld that confine the nunchucks in good focus within a fixed field of view, eliminating the need for constant adjustments and allowing for the automation of image capture. While these specialized flow cells are not commercially available due to the closure of businesses that once produced them, their previous existence demonstrates the feasibility of such technology. There is potential for these to be produced in-house, which could significantly lower the barriers to effective nunchuck use and imaging.

Bibliography

[1] J. C. Marini, S. D. Levene, D. M. Crothers, and P. T. Englund, *Bent helical structure in kinetoplast DNA*, *Proc. Natl. Acad. Sci. U.S.A.* **79** (1982), no. 24 7664–7668.

[2] H.-M. Wu and D. M. Crothers, *The locus of sequence-directed and protein-induced DNA bending*, *Nature* **308** (1984), no. 5959 509.

[3] N. Kaplan, I. K. Moore, Y. Fondufe-Mittendorf, A. J. Gossett, D. Tillo, Y. Field, E. M. Leproust, T. R. Hughes, J. D. Lieb, J. Widom, and Others, *The DNA-encoded nucleosome organization of a eukaryotic genome*, *Nature* **458** (2009), no. 7236 362.

[4] V. Iyer and K. Struhl, *Poly(dA:dT), a ubiquitous promoter element that stimulates transcription via its intrinsic DNA structure.*, *EMBO J.* **14** (1995), no. 11 2570–2579.

[5] K. K. Swinger and P. A. Rice, *IHF and HU: Flexible architects of bent DNA*, *Curr. Opin. Struct. Biol.* **14** (2004), no. 1 28–35.

[6] L. J. Maher, *Mechanisms of DNA bending*, *Curr. Opin. Chem. Biol.* **2** (Jan., 1998) 688–694.

[7] P. L. Privalov, A. I. Dragan, and C. Crane-Robinson, *The cost of DNA bending*, *Trends Biochem. Sci.* **34** (Sept., 2009) 464–470.

[8] B. M. Reinhard, S. Sheikholeslami, A. Mastroianni, a. P. Alivisatos, and J. Liphardt, *Use of plasmon coupling to reveal the dynamics of DNA bending and cleavage by single EcoRV restriction enzymes*, *Proc. Natl. Acad. Sci. U.S.A.* **104** (2007), no. 8 2667–2672.

[9] A. V. Pinheiro, D. Han, W. M. Shih, and H. Yan, *Challenges and opportunities for structural dna nanotechnology*, *Nature nanotechnology* **6** (2011), no. 12 763–772.

[10] Q. Zhang, Q. Jiang, N. Li, L. Dai, Q. Liu, L. Song, J. Wang, Y. Li, J. Tian, B. Ding, *et. al.*, *Dna origami as an in vivo drug delivery vehicle for cancer therapy*, *ACS nano* **8** (2014), no. 7 6633–6643.

[11] V. Linko, *At the dawn of applied dna nanotechnology*, 2020.

[12] A. Gopinath, E. Miyazono, A. Faraon, and P. W. Rothemund, *Engineering and mapping nanocavity emission via precision placement of dna origami*, Nature **535** (2016), no. 7612 401–405.

[13] A. E. Marras, L. Zhou, H.-J. Su, and C. E. Castro, *Programmable motion of dna origami mechanisms*, Proceedings of the National Academy of Sciences **112** (2015), no. 3 713–718.

[14] A. E. Marras, Z. Shi, M. G. Lindell III, R. A. Patton, C.-M. Huang, L. Zhou, H.-J. Su, G. Arya, and C. E. Castro, *Cation-activated avidity for rapid reconfiguration of dna nanodevices*, ACS nano **12** (2018), no. 9 9484–9494.

[15] J. S. Schreck, T. E. Ouldridge, F. Romano, A. A. Louis, and J. P. Doye, *Characterizing the bending and flexibility induced by bulges in DNA duplexes*, The Journal of chemical physics **142** (2015), no. 16 04B614_1.

[16] D. Lane, P. Prentki, and M. Chandler, *Use of gel retardation to analyze protein-nucleic acid interactions*, Microbiol. Mol. Biol. Rev. **56** (Dec., 1992) 509–528.

[17] D. M. Crothers, J. Drak, J. D. Kahn, and S. D. Levene in *Methods in Enzymology*, vol. 212, pp. 3–29. Elsevier, 1992.

[18] D. Macdonald, K. Herbert, X. Zhang, T. Polgruto, and P. Lu, *Solution structure of an A-tract DNA bend*, J. Mol. Biol. **306** (2001), no. 5 1081–1098.

[19] H. C. Nelson, J. T. Finch, B. F. Luisi, and A. Klug, *The structure of an oligo(dA)·oligo(dT) tract and its biological implications*, Nature **330** (1987), no. 6145 221.

[20] X. Shi, K. A. Beauchamp, P. B. Harbury, and D. Herschlag, *From a structural average to the conformational ensemble of a DNA bulge*, Proc. Natl. Acad. Sci. U.S.A. **111** (2014), no. 15 E1473–E1480.

[21] C. Rivetti, C. Walker, and C. Bustamante, *Polymer chain statistics and conformational analysis of DNA molecules with bends or sections of different flexibility*, J. Mol. Biol. **280** (July, 1998) 41–59.

[22] J. Bednar, P. Furrer, V. Katritch, A. Stasiak, J. Dubochet, and A. Stasiak, *Determination of DNA persistence length by cryo-electron microscopy. separation of the static and dynamic contributions to the apparent persistence length of DNA*, J. Mol. Biol. **254** (1995), no. 4 579–594.

[23] P. R. Heenan and T. T. Perkins, *Imaging DNA equilibrated onto mica in liquid using biochemically relevant deposition conditions*, *ACS Nano* **13** (Apr., 2019) 4220–4229.

[24] N. Destainville, M. Manghi, and J. Palmeri, *Microscopic mechanism for experimentally observed anomalous elasticity of DNA in two dimensions*, *Biophys. J.* **96** (2009), no. 11 4464–4469.

[25] A. K. Woźniak, G. F. Schröder, H. Grubmüller, C. A. Seidel, and F. Oesterhelt, *Single-molecule FRET measures bends and kinks in DNA*, *Proc. Natl. Acad. Sci. U.S.A.* **105** (2008), no. 47 18337–18342.

[26] A. Brunet, S. Chevalier, N. Destainville, M. Manghi, P. Rousseau, M. Salhi, L. Salomé, and C. Tardin, *Probing a label-free local bend in DNA by single molecule tethered particle motion*, *Nucleic Acids Res.* **43** (2015), no. 11 e72–e72.

[27] A. Basu, D. G. Bobrovnikov, Z. Qureshi, T. Kayikcioglu, T. T. Ngo, A. Ranjan, S. Eustermann, B. Cieza, M. T. Morgan, M. Hejna, *et. al.*, *Measuring dna mechanics on the genome scale*, *Nature* **589** (2021), no. 7842 462–467.

[28] T. E. Cloutier and J. Widom, *Spontaneous sharp bending of double-stranded dna*, *Molecular cell* **14** (2004), no. 3 355–362.

[29] P. A. Wiggins, T. Van Der Heijden, F. Moreno-Herrero, A. Spakowitz, R. Phillips, J. Widom, C. Dekker, and P. C. Nelson, *High flexibility of dna on short length scales probed by atomic force microscopy*, *Nature nanotechnology* **1** (2006), no. 2 137–141.

[30] R. Vafabakhsh and T. Ha, *Extreme bendability of dna less than 100 base pairs long revealed by single-molecule cyclization*, *Science* **337** (2012), no. 6098 1097–1101.

[31] M. Zoli, *J-factors of short dna molecules*, *The Journal of Chemical Physics* **144** (2016), no. 21 214104.

[32] M. Zoli, *Short dna persistence length in a mesoscopic helical model*, *EPL (Europhysics Letters)* **123** (2018), no. 6 68003.

[33] Q. Du, C. Smith, N. Shiffeldrim, M. Vologodskaia, and A. Vologodskii, *Cyclization of short dna fragments and bending fluctuations of the double helix*, *Proceedings of the National Academy of Sciences* **102** (2005), no. 15 5397–5402.

[34] A. Vologodskii, Q. Du, and M. D. Frank-Kamenetskii, *Bending of short dna helices*, *Artificial DNA: PNA & XNA* **4** no. 1 1–3.

[35] A. K. Mazur and M. Maaloum, *Dna flexibility on short length scales probed by atomic force microscopy*, *Physical review letters* **112** (2014), no. 6 068104.

[36] A. Basu, D. G. Bobrovnikov, and T. Ha, *Dna mechanics and its biological impact*, *Journal of molecular biology* **433** (2021), no. 6 166861.

[37] M. Kim, S. Bae, I. Oh, J. Yoo, and J. S. Kim, *Sequence-dependent twist-bend coupling in dna minicircles*, *Nanoscale* **13** (2021), no. 47 20186–20196.

[38] R. Saran, Y. Wang, and I. T. Li, *Mechanical flexibility of dna: a quintessential tool for dna nanotechnology*, *Sensors* **20** (2020), no. 24 7019.

[39] S. Yeou and N. K. Lee, *Single-molecule methods for investigating the double-stranded dna bendability*, *Molecules and cells* **45** (2022), no. 1 33.

[40] X. Cai, D. S. Arias, L. R. Velazquez, S. Vexler, A. L. Bevier, and D. K. Fygenson, *DNA nunchucks: Nanoinstrumentation for single-molecule measurement of stiffness and bending*, *Nano letters* **20** (2019), no. 2 1388–1395.

[41] A. M. Mohammed and R. Schulman, *Directing self-assembly of dna nanotubes using programmable seeds*, *Nano letters* **13** (2013), no. 9 4006–4013.

[42] A. Mohammed, L. Velazquez, a. Chisenhall, D. Schiffels, D. Fygenson, and R. Schulman, *Self-assembly of precisely defined DNA nanotube superstructures using DNA origami seeds*, *Nanoscale* **9** (2017), no. 2 522–526.

[43] P. W. K. Rothemund, *Folding DNA to create nanoscale shapes and patterns*, *Nature* **440** (Mar., 2006) 297–302.

[44] P. W. Rothemund, A. Ekani-Nkodo, N. Papadakis, A. Kumar, D. K. Fygenson, and E. Winfree, *Design and characterization of programmable DNA nanotubes*, *J. Am. Chem. Soc.* **126** (2004), no. 50 16344–16352.

[45] P. M. Nafisi, T. Aksel, and S. M. Douglas, *Construction of a novel phagemid to produce custom DNA origami scaffolds*, *Synth. Biol.* **3** (2018), no. 1 ysy015.

[46] É. Hegedüs, E. Kokai, A. Kotlyar, V. Dombradi, and G. Szabo, *Separation of 1–23-kb complementary dna strands by urea–agarose gel electrophoresis*, *Nucleic acids research* **37** (2009), no. 17 e112–e112.

[47] P. Šulc, F. Romano, T. E. Ouldridge, L. Rovigatti, J. P. Doye, and A. A. Louis, *Sequence-dependent thermodynamics of a coarse-grained dna model*, *The Journal of chemical physics* **137** (2012), no. 13.

[48] C. E. Aitken, R. A. Marshall, and J. D. Puglisi, *An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments*, *Biophysical journal* **94** (2008), no. 5 1826–1835.

[49] P. Mellado, A. Iniesta, F. Diaz, and J. G. De La Torre, *Diffusion coefficients of segmentally flexible macromolecules with two subunits: a study of broken rods*, *Biopolymers: Original Research on Biomolecules* **27** (1988), no. 11 1771–1786.

[50] A. Krizhevsky, I. Sutskever, and G. E. Hinton, *Imagenet classification with deep convolutional neural networks*, *Advances in neural information processing systems* **25** (2012).

[51] Y. Lu, B. Weers, and N. C. Stellwagen, *DNA persistence length revisited*, *Biopolymers* **61** (2002), no. 4 261–275.

[52] E. Stellwagen, J. P. Peters, L. J. Maher Iii, and N. C. Stellwagen, *DNA A-tracts are not curved in solutions containing high concentrations of monovalent cations*, *Biochemistry* **52** (2013), no. 24 4138–4148.

[53] J. S. Mitchell, J. Glowacki, A. E. Grandchamp, R. S. Manning, and J. H. Maddocks, *Sequence-dependent persistence lengths of DNA*, *J. Chem. Theory Comput.* **13** (2017), no. 4 1539–1555.

[54] E. N. Nikolova, G. D. Bascom, I. Andricioaei, and H. M. Al-Hashimi, *Probing sequence-specific DNA flexibility in A-tracts and pyrimidine-purine steps by nuclear magnetic resonance ^{13}C relaxation and molecular dynamics simulations*, *Biochemistry* **51** (Oct., 2012) 8654–8664.

[55] H. P. Lu, L. Xun, and X. S. Xie, *Single-molecule enzymatic dynamics*, *Science* **282** (1998), no. 5395 1877–1882.

[56] D. M. Rissin, H. H. Gorris, and D. R. Walt, *Distinct and long-lived activity states of single enzyme molecules*, *Journal of the American Chemical Society* **130** (2008), no. 15 5349–5353.

[57] C. K. Johnson, *Calmodulin, conformational states, and calcium signaling. a single-molecule perspective*, *Biochemistry* **45** (2006), no. 48 14233–14246.

[58] J. Shi, B. A. Palfey, J. Dertouzos, K. F. Jensen, A. Gafni, and D. Steel, *Multiple states of the tyr318leu mutant of dihydroorotate dehydrogenase revealed by single-molecule kinetics*, *Journal of the American Chemical Society* **126** (2004), no. 22 6914–6922.

[59] M. Connolly, A. Arra, V. Zvoda, P. J. Steinbach, P. A. Rice, and A. Ansari, *Static kinks or flexible hinges: multiple conformations of bent dna bound to integration host factor revealed by fluorescence lifetime measurements*, *The Journal of Physical Chemistry B* **122** (2018), no. 49 11519–11534.

[60] A. Basu, D. G. Bobrovnikov, B. Cieza, J. P. Arcon, Z. Qureshi, M. Orozco, and T. Ha, *Deciphering the mechanical code of the genome and epigenome*, *Nature structural & molecular biology* **29** (2022), no. 12 1178–1187.

[61] G. Park, M. K. Cho, and Y. Jung, *Sequence-dependent kink formation in short dna loops: Theory and molecular dynamics simulations*, Journal of Chemical Theory and Computation **17** (2021), no. 3 1308–1317.

[62] A. V. Drozdetski, A. Mukhopadhyay, and A. V. Onufriev, *Strongly bent double-stranded dna: Reconciling theory and experiment*, Frontiers in physics (2019) 195.

[63] F. Lankaš, R. Lavery, and J. H. Maddocks, *Kinking occurs during molecular dynamics simulations of small dna minicircles*, Structure **14** (2006), no. 10 1527–1534.

[64] C. Yuan, H. Chen, X. W. Lou, and L. A. Archer, *Dna bending stiffness on small length scales*, Physical review letters **100** (2008), no. 1 018102.

[65] R. A. Forties, R. Bundschuh, and M. G. Poirier, *The flexibility of locally melted DNA*, Nucleic Acids Res. **37** (Aug., 2009) 4580–4586.

[66] R. M. Harrison, F. Romano, T. E. Ouldridge, A. A. Louis, and J. P. Doye, *Identifying physical causes of apparent enhanced cyclization of short dna molecules with a coarse-grained model*, Journal of chemical theory and computation **15** (2019), no. 8 4660–4672.

[67] P. Cong and J. Yan, *Recent progress on the mechanics of sharply bent dna*, SCIENCE CHINA Physics, Mechanics & Astronomy **59** (2016), no. 8 1–8.

[68] L. Kozerski, A. Mazurek, R. Kawecki, W. Bocian, P. Krajewski, E. Bednarek, J. Sitkowski, M. Williamson, A. Moir, and P. Hansen, *A nicked duplex decamer dna with a peg6 tether*, Nucleic Acids Research **29** (2001), no. 5 1132–1143.

[69] C. Roll, C. Ketterlé, V. Faibis, G. V. Fazakerley, and Y. Boulard, *Conformations of nicked and gapped dna structures by nmr and molecular dynamic simulations in water*, Biochemistry **37** (1998), no. 12 4059–4070.

[70] E. Le Cam, F. Fack, J. Ménissier-de Murcia, J. A. Cognet, A. Barbin, V. Sarantoglou, B. Révet, E. Delain, and G. De Murcia, *Conformational analysis of a 139 base-pair dna fragment containing a single-stranded break and its interaction with human poly (adp-ribose) polymerase*, Journal of molecular biology **235** (1994), no. 3 1062–1071.

[71] C. Roll, C. Ketterlé, Y. Boulard, and G. Fazakerley, *Structure of oligonucleotides with either a strand break or a bulged nucleotide*, Biochimie **79** (1997), no. 9-10 593–597.

[72] E. Protozanova, P. Yakovchuk, and M. D. Frank-Kamenetskii, *Stacked–unstacked equilibrium at the nick site of dna*, Journal of molecular biology **342** (2004), no. 3 775–785.

[73] J. B. Mills, J. P. Cooper, and P. J. Hagerman, *Electrophoretic evidence that single-stranded regions of 1 or more nucleotides dramatically increase the flexibility of dna*, Biochemistry **33** (1994), no. 7 1797–1803.

[74] R. W. Roberts and D. M. Crothers, *Stability and properties of double and triple helices: dramatic effects of rna or dna backbone composition*, Science **258** (1992), no. 5087 1463–1466.

[75] P. P. Chan and P. M. Glazer, *Triplex dna: fundamentals, advances, and potential applications for gene therapy*, Journal of Molecular Medicine **75** (1997) 267–282.

[76] T. Zettl, X. Shi, S. Bonilla, S. M. Sedlak, J. Lipfert, and D. Herschlag, *The structural ensemble of a holliday junction determined by x-ray scattering interference*, Nucleic acids research **48** (2020), no. 14 8090–8098.

[77] C.-H. Hsieh and J. D. Griffith, *Deletions of bases in one strand of duplex dna, in contrast to single-base mismatches, produce highly kinked molecules: possible relevance to the folding of single-stranded nucleic acids.*, Proceedings of the National Academy of Sciences **86** (1989), no. 13 4833–4837.

[78] M. A. Rosen, D. Live, and D. J. Patel, *Comparative nmr study of an-bulge loops in dna duplexes: intrahelical stacking of a, aa, and aaa bulge loops*, Biochemistry **31** (1992), no. 16 4004–4014.

[79] Y. H. Wang and J. Griffith, *Effects of bulge composition and flanking sequence on the kinking of dna by bulged bases*, Biochemistry **30** (1991), no. 5 1358–1363.

[80] F. Stühmeier, A. Hillisch, R. M. Clegg, and S. Diekmann, *Fluorescence energy transfer analysis of dna structures containing several bulges and their interaction with cap*, Journal of molecular biology **302** (2000), no. 5 1081–1100.

[81] X. Shi, K. A. Beauchamp, P. B. Harbury, and D. Herschlag, *From a structural average to the conformational ensemble of a DNA bulge*, Proceedings of the National Academy of Sciences **111** (2014), no. 15 E1473–E1480.

[82] T. Tran and B. Cannon, *Differential effects of strand asymmetry on the energetics and structural flexibility of dna internal loops*, Biochemistry **56** (2017), no. 49 6448–6459.

[83] T. Paul, S. C. Bera, and P. P. Mishra, *Direct observation of breathing dynamics at the mismatch induced dna bubble with nanometre accuracy: a smfret study*, Nanoscale **9** (2017), no. 18 5835–5842.

[84] K. A. Schallhorn, K. O. Freedman, J. M. Moore, J. Lin, and P. C. Ke, *Single-molecule dna flexibility in the presence of base-pair mismatch*, Applied Physics Letters **87** (2005), no. 3.

[85] C. Yuan, E. Rhoades, X. W. Lou, and L. A. Archer, *Spontaneous sharp bending of dna: role of melting bubbles*, Nucleic acids research **34** (2006), no. 16 4554–4560.

[86] L. Jaeger and A. Chworos, *The architectonics of programmable rna and dna nanostructures*, Current opinion in structural biology **16** (2006), no. 4 531–543.

[87] D. Lilley, *Kinking of dna and rna by base bulges.*, Proceedings of the National Academy of Sciences **92** (1995), no. 16 7140–7142.

[88] L. Huang and D. M. Lilley, *A quasi-cyclic rna nano-scale molecular object constructed using kink turns*, Nanoscale **8** (2016), no. 33 15189–15195.

[89] C. Laing and T. Schlick, *Analysis of four-way junctions in rna structures*, Journal of molecular biology **390** (2009), no. 3 547–559.

[90] S. Hohng, T. J. Wilson, E. Tan, R. M. Clegg, D. M. Lilley, and T. Ha, *Conformational flexibility of four-way junctions in rna*, Journal of molecular biology **336** (2004), no. 1 69–79.

[91] K. Réblová, N. Spackova, J. E. Sponer, J. Koca, and J. Sponer, *Molecular dynamics simulations of rna kissing–loop motifs reveal structural dynamics and formation of cation-binding pockets*, Nucleic acids research **31** (2003), no. 23 6942–6952.

[92] Y. G. Yingling and B. A. Shapiro, *Computational design of an rna hexagonal nanoring and an rna nanotube*, Nano letters **7** (2007), no. 8 2328–2334.

[93] S. M. Dibrov, J. McLean, J. Parsons, and T. Hermann, *Self-assembling rna square*, Proceedings of the National Academy of Sciences **108** (2011), no. 16 6405–6408.

[94] C. Geary, A. Chworos, and L. Jaeger, *Promoting rna helical stacking via a-minor junctions*, Nucleic acids research **39** (2011), no. 3 1066–1080.

[95] M. N. Bui, M. B. Johnson, M. Viard, E. Satterwhite, A. N. Martins, Z. Li, I. Marriott, K. A. Afonin, and E. F. Khisamutdinov, *Versatile rna tetra-u helix linking motif as a toolkit for nucleic acid nanotechnology*, in *Therapeutic RNA Nanotechnology*, pp. 223–246. Jenny Stanford Publishing, 2021.

[96] T. Shibata, Y. Fujita, H. Ohno, Y. Suzuki, K. Hayashi, K. R. Komatsu, S. Kawasaki, K. Hidaka, S. Yonehara, H. Sugiyama, *et. al.*, *Protein-driven rna nanostructured devices that function in vitro and control mammalian cell fate*, Nature communications **8** (2017), no. 1 540.

[97] H. Saito, T. Kobayashi, T. Hara, Y. Fujita, K. Hayashi, R. Furushima, and T. Inoue, *Synthetic translational regulation by an l7ae–kink-turn rnp switch*, Nature chemical biology **6** (2010), no. 1 71–78.

[98] P. A. Rice, *Making dna do a u-turn: Ihf and related proteins, Current opinion in structural biology* **7** (1997), no. 1 86–93.

[99] Y. Velmurugu, P. Vivas, M. Connolly, S. V. Kuznetsov, P. A. Rice, and A. Ansari, *Two-step interrogation then recognition of dna binding site by integration host factor: an architectural dna-bending protein, Nucleic acids research* **46** (2018), no. 4 1741–1755.

[100] J. Zhang, M. J. McCauley, L. J. Maher III, M. C. Williams, and N. Israeloff, *Mechanism of dna flexibility enhancement by hmgb proteins, Nucleic acids research* **37** (2009), no. 4 1107–1114.

[101] Z. Chen, H. Yang, and N. P. Pavletich, *Mechanism of homologous recombination from the reca–ssdna/dsdna structures, Nature* **453** (2008), no. 7194 489–494.

[102] D. Schiffels, V. A. Szalai, and J. A. Liddle, *Molecular precision at micrometer length scales: hierarchical assembly of dna–protein nanostructures, ACS nano* **11** (2017), no. 7 6623–6629.