

Software Tools for Design, Simulation, and Characterization of DNA and RNA Nanostructures

by

Erik Poppleton

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved June 2022 by the
Graduate Supervisory Committee:

Petr Šulc, Chair
Hao Yan
Stephanie Forrest
Nicholas Stephanopoulos

ARIZONA STATE UNIVERSITY

August 2022

ABSTRACT

Nucleic acid nanotechnology is a field of nanoscale engineering where the sequences of deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) molecules are carefully designed to create self-assembled nanostructures with higher spatial resolution than is available to top-down fabrication methods. In the 40 year history of the field, the structures created have scaled from small tile-like structures constructed from a few hundred individual nucleotides to micron-scale structures assembled from millions of nucleotides using the technique of "DNA origami". One of the key drivers of advancement in any modern engineering field is the parallel development of software which facilitates the design of components and performs *in silico* simulation of the target structure to determine its structural properties, dynamic behavior, and identify defects. For nucleic acid nanotechnology, the design software CaDNAno and simulation software oxDNA are the most popular choices for design and simulation, respectively. In this dissertation I will present my work on the oxDNA software ecosystem, including an analysis toolkit, a web-based graphical interface, and a new molecular visualization tool which doubles as a free-form design editor that covers some of the weaknesses of CaDNAno's lattice-based design paradigm. Finally, as a demonstration of the utility of these new tools I show oxDNA simulation and subsequent analysis of a nanoscale leaf-spring engine capable of converting chemical energy into dynamic motion. OxDNA simulations were used to investigate the effects of design choices on the behavior of the system and rationalize experimental results.

DEDICATION

To Phillip Poppleton and Wayne Hentschel for introducing me and so many others to the wonders of the natural world.

ACKNOWLEDGMENTS

Getting a Ph.D is a long journey, and you never know where your curiosity is going to take you. I came to ASU to study bioremediation, imagine my surprise five years later to be writing a dissertation on the software tools I wrote for molecular dynamics simulations of nanostructures. This awesome journey would not have been possible without the support, and guidance of my advisors, colleagues, family, and friends.

First, I want to thank my advisors, Professors Petr Šulc and Hao Yan. I first met both of them when I emailed Hao about potentially doing a rotation in his lab. It just so happened that the day I met with him was also the day that Petr was there to sign his contract. That chance meeting, Petr's simple question, "Can you code?" and my not-entirely-confident response touched off an incredibly fruitful collaboration which resulted in more than 10 publications and a software framework which is now used around the world. Hao is a quiet, yet powerful presence in the DNA nanotechnology field, he leads a lab filled with excellent and dedicated students with a light touch that encourages us to develop solutions collaboratively. Petr, meanwhile, is a jovial presence, alternating between cracking jokes and dropping yet another grant application in need of feedback on my desk. I am deeply indebted to those grants for funding my entire Ph.D, and to all the banter over lunches of hot chicken for keeping me sane. Petr, I hope you get that Nobel Prize for solving crystal lattice design so you can spend the rest of your days selling "geometric water" to the hippies in Sedona.

A huge thank you to my committee members and my collaborators on this journey. To Prof. Nicholas Stephanopoulos, your excitement about new ways to build nanostructures is infectious. To Prof. Stephanie Forrest, your class was so inspiring, looking at biology from a computing perspective was invaluable in helping me understand algorithms. To Prof. Fei Zhang, thank you for your mentorship during my first year in the lab. Your talk on difficult projects at your final lab meeting has meant so much to me and has kept me going through many late nights. To Prof. Lorenzo Rovigatti, I and the rest of the field cannot thank you enough for your continued work on oxDNA, it has been an honor to work alongside you to improve the software. I want to thank my collaborators, Profs. Giulia Pedrielli, Dimitri Bertsekas, Michael Famulok and Yonggang Ke for expanding my research horizons and inviting me to work on many awesome projects. And to Prof. Anca Delgado, thank you for welcoming me when I came to ASU. I learned so much about research and project planning from you and though I didn't end up joining your lab, I will always treasure the time I spent there.

None of what you see in this dissertation would have happened without Dr. Michael Matthies. He is the mad scientist who wrote the prototype for what would become oxView in a single night. His wide-eyed excitement for making better tools was inspiring, and though I often had to play the realist to his idealist, I couldn't have asked for a better collaborator to spend my Ph.D with. The rest of the Šulc lab as well, you have all been

fantastic collaborators and friends. Hao Liu, Jonah Procyk, Joakim Bohlin, Matthew Sample, Lanshen Zhao and Joshua Evans, thank you for all the ideas you've shared and code you've tested.

And to the members of the Yan Lab, past and present. I would first like to thank Dr. Guangbao Yao for being my mentor. I would also like to specifically thank Kristen Lee, Dr. Xiaodong Qi, Liangxiao Chen and Lu Yu for helping to manage the lab. And everybody else, Dr. Swarup Dey, Dr. Raghu Pradeep, Dr. Xu Zhou, Dr. Shouxing Jiang, Dr. Chad Simmons, Dr. Minghui Liu, Dr. Yang Xu, Leeza Abraham, Yue Tang, Abhay Prasad, Deeksha, Rong Zheng and Xinyi Tu, thank you for all your advice and patience as I bounced between labwork and simulation and didn't always remember where we kept the acetic acid.

I want to thank my undergraduate mentees, Teddy Yellowman, Shuchi Sharma, Roger Romero, Jacob Baca, Kai Yin, Jeffrey Luo, Christopher Persson, and Aatmik Mallya. It was fantastic working with all of you.

Next, I want to thank the three high school science teachers who started me down this journey: Wayne Hentschel, Dr. Gary Ayton, and Dr. Stephanie Atherton. You prepared me for my career in science incredibly well. I specifically want to thank Dr. Ayton for first introducing me to molecular dynamics simulations.

Now, for the part where I get to thank all the friends and family who supported me through this journey. First, I would like to thank my family, Ken (dad), Staci (step-mom), Jodi (sister), Lisa (mom), Jim (step-dad) for their endless support and all the lessons big and small they've taught me over the years.

I also want to thank the friends from far away who were still always there for me. To the Salt Lake squad: Katie, Angel, Nicole, Jack, Bryan, Pavitra, Sharmista, Jeremey, and Natalia, thank you for always being the home I could return to. To my late-night gamers: Elizabeth, Will, and Emily, thank you for being my constant companions and greatest emotional supports. And to the one and only Lounge Crew: Amy, Rhiannon, Nolan, Ben, Aahlad, Jei-Jei, Shane, Ruth, Ashley, Stephen, Saurav, Zabir, Justice, Phoebe, Jenni, and Dr. Ritah for being my 24/7 source of discourse, discord, and wisdom.

I want to thank my fellow and former Phonecians, Dr. Valerie, Dr. Gracie, Chinami, George, Cassie, Linnea, Sean, Eliot, and Dr. Brad for all the hiking, art, and cooking. I also want to specifically thank the extended family of the Delgado lab: Moni, Aide, Sayalee, Srivatsan, Justin, Skanda, Kelise, Ashley, Rupinder, and Felipe for being there for me since the very beginning and making Biodesign feel like home.

And finally, I want to acknowledge the microbe in the room: SARS-CoV-2. Living, let alone finishing a Ph.D, in the time of Covid should have been hard, but I was blessed with the best pandemic pod: Jen, Liza, Dani, Matt and Elliot, you made the last two years way better than they had any right to be, I'm not sure what I would have done without you. Also, thank you for letting me be the pseudo-father to your weird dogs: Daisy, Hadron, Tex, Honeybee, and Pippin; they helped me so much.

TABLE OF CONTENTS

	Page
LIST OF TABLES	viii
LIST OF FIGURES	ix
CHAPTER	
1 INTRODUCTION	1
1.1 Engineering at the Nanoscale	1
1.1.1 The Promise and Challenges of Nanotechnology	1
1.1.2 The Atoms that Make up Everything	4
1.1.3 DNA Nanotechnology	5
1.1.4 RNA Nanotechnology	8
1.1.5 DNA and RNA Nanostructure Design Software	10
1.2 The oxDNA and oxRNA Models	13
1.2.1 Molecular Models: Applications and Methods	13
1.2.2 Coarse-Graining DNA, RNA and Proteins	17
1.2.3 Limitations of the Model	20
1.2.4 Choosing the Correct Simulation Methods	21
1.3 Motivation For a Better oxDNA Software Environment	22
1.4 A Co-author's Introduction to Chapter 4	23
1.5 Full List of Publications	24
2 SOFTWARE TOOLS FOR PREPARATION, VISUALIZATION, AND ANALYSIS OF OXDNA/RNA SIMU- LATIONS	26
2.1 Introduction	26
2.2 Materials and Methods	29
2.2.1 System and Software Requirements	29
2.2.2 Simulation Details	29
2.3 Results	30
2.3.1 OxView - Web Browser Visualization, Analysis and Editing of Nanostructures	30
2.3.1.1 Implementation Details	32
2.3.1.2 Data Overlays in oxView	32
2.3.1.3 Relaxing Structures Using Rigid Body Dynamics	33

CHAPTER	Page
2.3.2 General–Purpose Analysis Tools	34
2.3.2.1 Mean Structure Determination and RMSFs	35
2.3.2.2 Geometric Parameters: Interduplex Angles and Distances	38
2.3.2.3 Base Pair Occupancy	40
2.3.2.4 Principal Component Analysis of Nanostructure Motion Modes	41
2.3.2.5 Unsupervised Clustering of Configurations Encountered in Simulation	42
2.3.2.6 Other Utilities	43
2.4 Discussion	45
2.5 Data Availability	47
2.6 Acknowledgements	47
2.7 Funding	47
2.7.1 Conflict of Interest Statement.	47
3 ONLINE RESOURCES FOR CHARACTERIZING DNA AND RNA NANOSTRUCTURE DESIGNS	48
3.1 Introduction	48
3.2 Methods	50
3.2.1 Server Data Processing	50
3.2.2 Software	53
3.3 Results and Discussion	55
3.3.1 Input Files for the Server	55
3.3.2 Server Output Result	56
3.4 Conclusion	57
3.5 Acknowledgements	58
3.6 Funding	58
4 USING OXDNA SIMULATIONS TO CHARACTERIZE THE EFFECTS OF DESIGN CHOICES ON A LEAF– SPRING NANOENGINE	59
4.1 Introduction	59
4.2 Design of the Nanoengine	60
4.3 Molecular Dynamics Simulations	61
4.4 Driver–Follower Experiments	65
4.5 Conclusion	68

CHAPTER	Page
4.6 Author Contributions	70
4.7 Additional Information	71
4.8 Acknowledgements	71
5 CONCLUSION AND FUTURE OUTLOOKS	72
REFERENCES	76
APPENDIX	
A SUPPLEMENTARY MATERIAL FOR CHAPTER 2	91
B SUPPLEMENTARY MATERIAL FOR CHAPTER 4	97
C PLEASE SEE ATTACHED FILES	108
D PERMISSIONS FOR MANUSCRIPT RE-USE	110
E PERMISSIONS FOR ADAPTED FIGURES	112

LIST OF TABLES

Table	Page
1. Browser Compatibility	58

LIST OF FIGURES

Figure	Page
1. Examples of Functional Nanomaterials	2
2. Biological Nanomaterials	6
3. The Evolution of DNA Nanotechnology	8
4. DNA Design Interfaces	12
5. Structure of DNA and RNA Helices	18
6. The oxDNA Model	28
7. Screenshots from Usage of oxView	30
8. Rigid-Body Dynamics of Clusters	33
9. Mean Structures and RMSF	36
10. Improving Mean Structures of Flexible Designs	37
11. Centroid Structure and Mean Computed via Multidimensional Scaling	39
12. Bond Occupancy of an RNA Tile	41
13. Principal Component Analysis of a Holliday junction	42
14. Unsupervised Clustering to Isolate Isomers of an RNA Tile	44
15. The oxDNA Model	49
16. Workflow of the oxDNA.org Pipeline	50
17. Submitting a Job on oxDNA.org	54
18. Analyzing a Completed Job	56
19. Design and Dimensions of the DNA Leaf-Spring Nanoengine	63
20. Coarse-Grained Simulation of Nanoengine Designs	66
21. Driving a Passive Follower Unit	70
22. oxDNA Ecosystem	73
23. Building the Leaf-Spring Nanoengine in oxView	74
A1. Configuration Positions in Principal Component Space	94
A2. Comparing Angles in Wireframe Lattices	95
A3. Distance Between Origami Units of TMF	96
A4. 3D Printed Holliday Junction Exported from oxView	96
B1. Additional Design Features of the Leaf-Spring NE	101
B2. Assembly of the NE and AFM Analysis of the Geometry of the NE	103

Figure	Page
B3. Comparison of Hinge Properties of Simulations run at 23° and 37° C	105
B4. Driver Follower Experiments	107
D1. License Information for Chapter 2	111
D2. License Information for Chapter 3	111
E1. Permissions for Figure 1	113
E2. Permissions for Figure 1	114
E3. Permissions for Figure 1	114
E4. Permissions for Figure 2	115
E5. Permissions for Figure 2	115
E6. Permissions for Figure 2	116
E7. Permissions for Figure 3	116
E8. Permissions for Figure 3	117
E9. Permissions for Figure 3	117

Chapter 1

INTRODUCTION

1.1 Engineering at the Nanoscale

1.1.1 The Promise and Challenges of Nanotechnology

Human achievement has long been measured by the things we build, from the first stone tools and early irrigation to modern wonders like smartphones, particle accelerators and high-speed rail systems. Every engineering project requires an intimate understanding of the materials used in construction. Prehistoric peoples used obsidian, flint and jasper for their stone tools due to their ability to hold a cutting edge and the ease of shaping them through knapping. Though they did not understand the molecular structures that impart these features, their macroscale usefulness was clear to stone age engineers. Today, with modern chemical and physical techniques, we can peer down to the scale of individual atoms and marvel at the diverse and complex structures that make up the fundamental structures of all things. The obvious question then arises — how can we build and control at the molecular scale to create novel materials and functionalities?

“Nanotechnology”, the discipline of building at the nanometer scale is an engineering challenge. Recent developments in the field have given us such technological marvels as the cheap, miniaturized electronics that have led to a proliferation of “smart” everything, liposome-based drug delivery platforms which are core to modern cancer therapies and the delivery vehicle in mRNA vaccines, extremely high surface-area catalysts used in chemical production, and the food we consume often contains additives which are nanostructured to enhance delivery efficiency and come in packaging made of nanocomposite materials which are frequently used in packaging to maintain freshness and impart diagnostic function in food and agricultural products¹ (Figure 1).

Many people point to Richard Feynman’s 1959 talk “There’s Plenty of Room at the Bottom”⁸ as the guiding inspiration for the field. In it he identified the potential of the field — namely the ability to craft custom molecules and the incredible information density which could be achieved by writing and reading information at a molecular scale. Things certainly have shrunk a lot since 1959; the tools and methods for engineering at

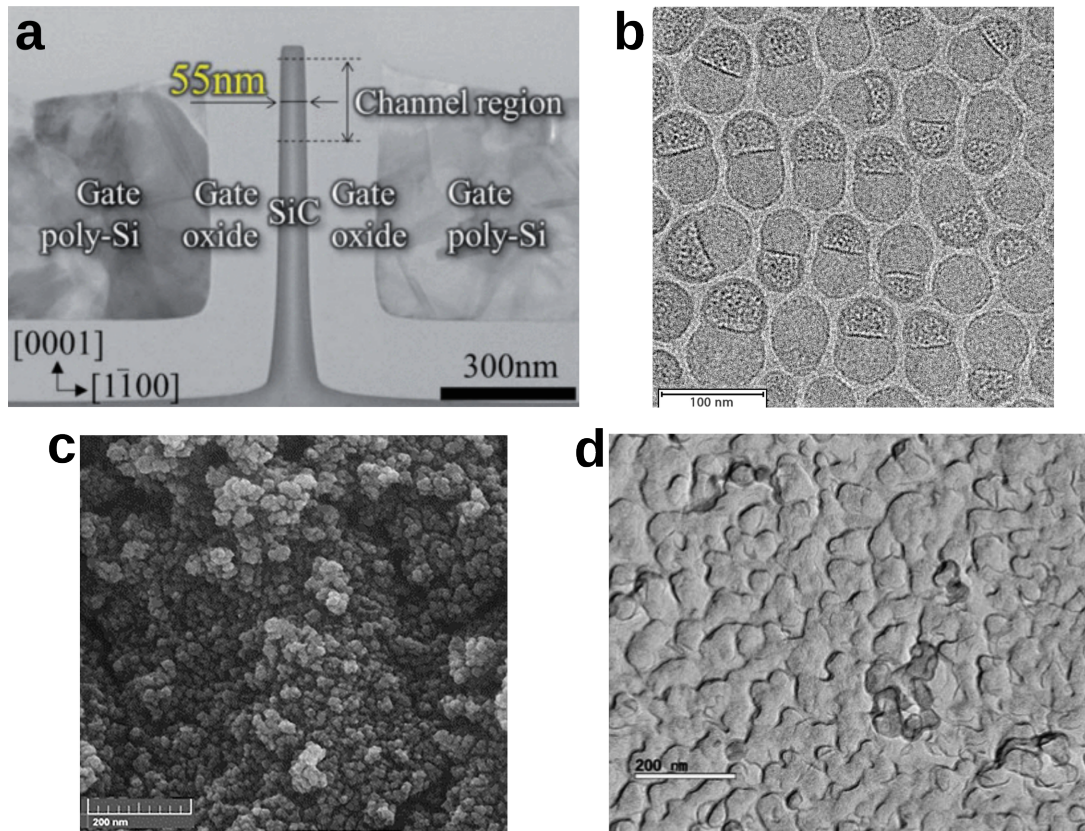


Figure 1. **Examples of functional nanomaterials.** (a) TEM image of a low-pitch field-effect transistor. Transistors are the basic information processing unit in all digital computers. (b) TEM image of Moderna's lipid nanoparticle-based mRNA vaccine against SARS-CoV-2. The lipid nanoparticles facilitate cellular uptake allowing production of the antigen from the mRNA cargo. (c) SEM image of a Pd/Fe₃O₄ catalyst used in biphenol production. Note the large surface area created by aggregation of nanoparticles. (d) TEM image of a β -carotene nanoemulsion. This type of emulsion are used as a food additive to increase delivery of lipophilic food additives². Images re-used with permission (a) Copyright 2020 IEEE³. (b) Used under a creative commons license. Obtained from⁴ which adapted the figure from⁵. (c) Copyright Elsevier 2019⁶, (d) Copyright Elsevier 2008⁷

the nanometer and sub-nanometer scale we have today are technological marvels in their own right, but yet many challenges still remain in building down at the bottom.

It can be difficult to conceptualize how small a nanometer is. One of the thinnest things we experience on the daily basis, our own hair, is on average, $75 \mu\text{m}$, or 7500 nm in diameter. This is Feynman's main point, there isn't just space to engineer below the realms of our human perception, but there's *vast amounts* of space to engineer down there. Building at bottom of matter isn't a trivial proposition, however. At the nanoscale, traditional tools we use for construction are no longer viable:

1. Grabbing and manipulating an individual component is extremely difficult. There are a number of tech-

niques including optical tweezers⁹, atomic force microscopy (AFM)¹⁰, and microlithography (electron beam or photolithography)¹¹ which have been used to successfully construct nanodevices. However, these techniques generally require extremely expensive and sensitive equipment, and the throughput is relatively low (the major exception are semiconductor manufactures which have developed robust methods to mass produce computer chips using photolithography at multibillion dollar fabrication facilities¹²)

2. Nanoparticles, especially those biological in origin or used in biotechnological applications exist in an aqueous environment. At the nanoscale, the physics of fluids are totally different than at the macroscale¹³. One of the most striking differences between our day-to-day experience with water and what micro- and nanoscale objects experience is related to inertia, namely that for small objects, friction from the surrounding medium dominates inertial forces. This ratio between friction and inertia, the "Reynolds number", affects what types of motion are relevant for control of an object. Only cyclic motions, rather than linear motions are capable of driving overall translation for very small objects¹⁴.
3. In addition to the problem of drag on directed motion introduced by the solvent, at the nanoscale thermally-driven motion, especially collisions with solvent molecules provide constant dynamics which affect material properties¹⁵. At the atomic scale, each atom is fluctuating with respect to its neighbors due to thermal fluctuations, making precise positioning impossible¹⁶. Many nanoscale objects of relevance, especially those discussed here are polymer chains, which due to their linear structure experience large conformational changes in response to relatively small angular displacements of individual elements. Due to constant collisions with fluid, flexible polymers undergo constant fluctuation and subtle changes in local interactions between individual monomers and between monomers and solvent can lead to large changes in global conformation dynamics¹⁷.
4. Particle aggregation is also a serious consideration when engineering at the nanoscale. Aggregation reduces surface area, which can be detrimental for nanoparticle function, especially for those used in surface-based catalysis¹⁸ or precise structural engineering¹⁹. The reasons for this behavior are complex, including assembly defects, Van der Waals forces between surface atoms, and entropically-driven hydrophobic interactions arising from displacement of ordered solvent molecules²⁰. Controlling these behaviors requires careful consideration of nanoparticle concentration, surface design, and solvent pH.

As scientists and engineers learn to work at the nanoscale, one of the key determinants of success will be computational technologies which allow for *in silico* design and prediction of nanoparticle structure and behavior. Because of the differences in behavior and manufacturing detailed in the previous section, these

tools must be specifically tailored for the nanoscopic regime. In any engineering discipline there are two invaluable types of software: Computer Aided Design (CAD) tools and simulation engines. CAD tools let users build using higher-level abstractions and re-use a common toolkits of parts, allowing efficient fabrication and standardization. Simulations are important for prototyping structures, allowing engineers to predict the behaviors and verify the structural integrity of designs prior to fabrication which can be expensive in both time and materials. Simulations allow parallelization of design iteration and can provide a level of resolution through the modeling of physical behaviors that cannot be achieved through experimental characterization.

Currently, with nanoscale science still in its infancy, the tools used, particular for design, are still in flux and undergoing constant development. There have been some attempts to create multi-material, generalized CAD tools^{21,22}, however none have gained common adoption due to lack of utility and poor user interfaces. As nanotechnology as a field advances, the need for more advanced and user-friendly design and simulation options will grow as we learn how to create more complex and feature-rich nanostructures.

1.1.2 The Atoms that Make up Everything

When designing nanomaterials, the constituent atoms define both the end function and the construction process. Nanomaterials can be broken down into four broad categories: information processing (electronic and optical materials), catalysis, structural, and biomedical. The most advanced applications are currently in the semiconductor transistor space where advanced optical lithography fabrication is able to create multi-layered structures out of silicon and metal oxides at the tens of nanometer scale. Transistors are able to selectively permit and deny the flow of electrons, allowing information processing. Similarly, there is active research into the creation of plasmonic²³ and photonic²⁴ materials which are able to direct electromagnetic fields and photons in predefined manners and have myriad applications ranging from electronic-like optical transistors²⁵, to surface plasmon resonance spectroscopy²⁶ which has become a standard method for detecting molecular interactions in biological research. Catalytic nanomaterials take advantage of the high surface area created by nanoscale structures to increase the number of catalytic units significantly above what would be available to macroscale-structured materials. For engineered applications, these are most often metallic structures²⁷, however it can be argued that the enzymes that perform catalysis in living systems are also nanomaterials, with angstrom-scale placement of electrophilic and nucleophilic groups to catalyze specific organic reactions. On the structural side, polymer hydrogels, cytoskeletal mimetics, and enzyme scaffolds are all types of engineered nanoscale materials which provide support ordering for other nanoscale components.

As previously mentioned regarding proteins, the internal mechanisms of cells are naturally occurring active nanostructures which fit into all of the aforementioned categories. DNA, RNA and associated transcription factors perform computing, enzymes perform catalysis, actin, tubulin and membranes with associated proteins create structural motifs which range in scale from nanoscale pits and pores to microscale cellular morphology. Engineered nanomaterials which can interface with biology are of high research interest. The functions of these materials range from basic biophysics, to diagnostics, to drug delivery to prosthetics. One of the most impressive features of biological nanostructures is that they are entirely self-assembled. Unlike modern semiconductors which are built using the “top-down” method of optical lithography, all the structures observed in biological systems assemble from the “bottom-up,” emerging from interactions between relatively few types of atoms. One of the key features in biology’s ability to create highly diverse structures is the use of polymers which fold into a defined 3D structure. Polymers have multiple advantages over monomeric units when it comes to self-assembling structures: they overcome the “diffusion problem” which arises when low concentrations of molecules need to be able to find each other in order to begin to assemble by having very high local concentration of the monomeric units and by facilitating diffusion of other molecules along the contour length of the polymer²⁸ allowing interaction and assembly of both intra- and intermolecular complexes; they also allow the combinatoric assembly of relatively few types of monomers into highly diverse structures and data-containing motifs. This “aperiodic crystal” as postulated by Schrödinger²⁹ which is simultaneously ordered — it has a rigid sequence of units drawn from a limited pool of constituents, but does not have a repeating structure is common to all three major groups of biological macromolecules, nucleic acids, amino acids, and polysaccharides.

1.1.3 DNA Nanotechnology

Biologists are very familiar with thinking about DNA as the carrier of genetic information within living systems. All the information required to build a living organism can be found in the sequence of adenine (A), thymine (T), guanine (G) and cytosine (C) bases (for a human, that’s a bit more than 3 billion base pairs³³). However, one of the other features of DNA is its consistent structure: the B-form double helix³⁴. The double helix forms from the interplay of molecular structures within the DNA molecule. The pi-pi stacking between neighboring bases forces a shorter inter-base distance than would be ideal for the more extended sugar-

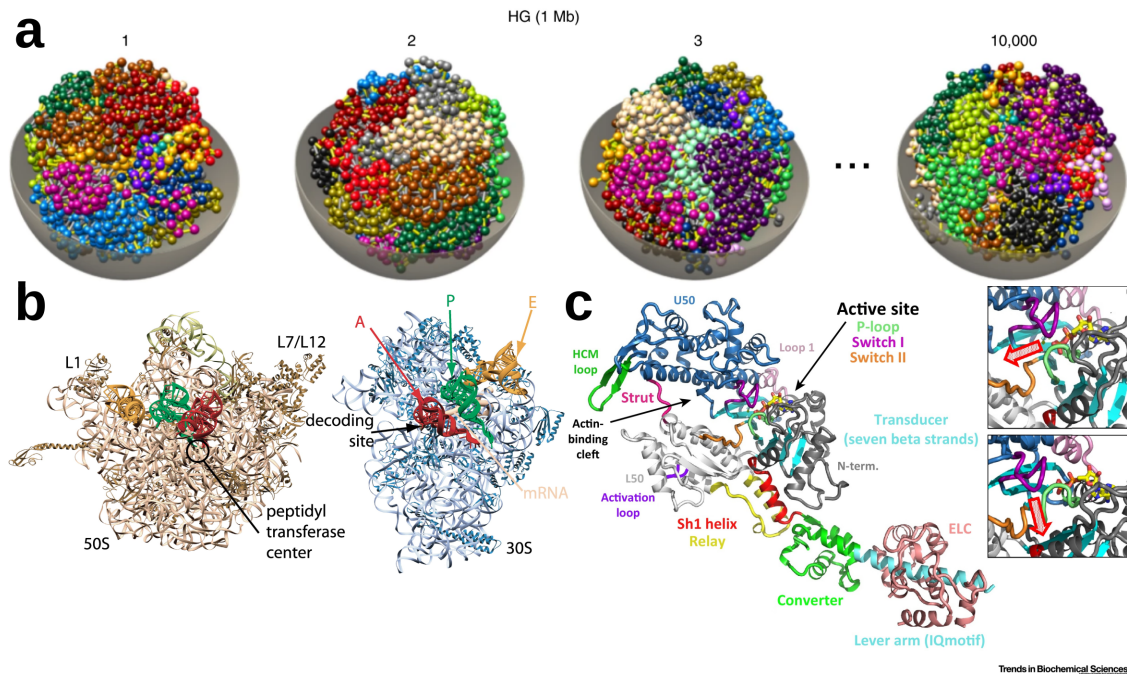


Figure 2. **Cellular components are nanomachines** (a) Chromosome organization within the nucleus imaged by GPSeq and Hi-C³⁰. Each colored strand represents a chromosome and each bead corresponds to 1 million base pairs. DNA is an extremely dense information storage medium which undergoes constant regulation, remodeling, and copying by associated proteins and RNA. (b) The structure of the two ribosomal subunits. The ribosome is a nanoscale RNA/protein structure which catalyzes the peptidyl transferase reaction responsible for building all proteins. (c) Protein structure of the myosin motor domain. Myosin proteins “walk” along actin fillaments inside cells, delivering cargo and moving actin fillaments. The collective action of millions of myosin proteins gives rise to muscle contractions. Images re-used with permission (a) Copyright 2020 Springer Nature³⁰. (b) Copyright 2002 Elsevier³¹. (c) Copyright 2022 Elsevier³²

phosphate covalent backbone. This causes the two molecules to wind around each other in a right-handed helix with a pitch of 3.57 nm (Figure 5a).

In 1982, Nadrian Seeman identified that through careful design and synthesis of DNA sequences, it would be possible to coax DNA to self-assemble into arbitrary structures composed of immobile junctions³⁵. In biology, DNA forms “Holliday junctions” during genetic recombination events³⁶. These junctions eventually resolve back to linear duplexes through strand displacement as both arms of the junction share the same sequence. Seeman realized that by synthesizing oligonucleotide strands which would form junctions with distinct sequences on each arm, they would become immobile; furthermore, by chaining multiple of these junctions together, rigid structures could be created out of DNA. The original goal of this method was to crystallize the DNA lattices and use them as a substrate for difficult-to-crystallize proteins to aid in solving their structure crystallographically. Though the ability to tether proteins with sufficient rigidity in a DNA lattice has not yet been realized, DNA nanotechnology has become one of the most popular methods for assembling

nanostructures because of the relative inexpensive nature of the components (custom DNA oligo synthesis is down to cents per base[37]), the robustness of the assembly methods, the ability to position other components with ~ 2 nm resolution and the intrinsic biocompatibility of using DNA as a material.

After the initial idea and realization of assembly using DNA tiles³⁸, the field mostly focused on creating small polyhedra³⁹⁻⁴¹ and unbounded crystalline assemblies⁴², primarily focused on creating ordered arrays of substrates⁴³, and massively parallel automata-based computing⁴⁴⁻⁴⁶ which give more structure to the "molecular computing" idea developed by Leonard Adleman in 1994⁴⁷. The next major leap in the field came in 2006 when Paul Rothemund developed the technique of "DNA origami" wherein a single long strand of DNA (often a M13 bacteriophage genome due to the ease of culturing them for large quantities of single-stranded DNA) is folded by numerous shorter "staple" oligonucleotide strands a few dozen nucleotides in length which bridge 2-4 domains on the M13 scaffold⁴⁶. This allowed DNA nanostructures to scale to significantly larger 2D^{48,49} and 3D⁵⁰⁻⁵² fully-addressible structures, which when carefully designed to assemble into complexes, are able to create micrometer-scale structures which are still able to precisely position guest molecules with the same nanometer precision of the smaller DNA sequences⁵³⁻⁵⁵. In addition by combining structural DNA nanotechnology with molecular computing, it is possible to create dynamic structures which react to the environment⁵⁶.

The maturation of the field has led to a rapid increase in the number of applications of DNA origami. Much effort in the field has been put into developing DNA nanostructure-based therapeutics⁵⁷, including precision drug delivery platforms⁵⁸⁻⁶⁰, immunotherapy⁶¹⁻⁶³, and gene therapy (particularly silencing using antisense oligonucleotides)⁶⁴⁻⁶⁶. Nanoelectronics assembled from DNA are currently at a less mature state than biomedical applications, however, there has been progress using DNA structures to position carbon nanotubes in the structure of a field-effect transistor with a pitch 3-4x smaller than the finest structures achievable by top-down photolithography^{67,68}. Some of the most advanced usage of DNA nanostructures is in the area of basic biological research which does not require overcoming regulatory or scaled-up manufacturing hurdles in order to reach the application stage. These include measurements of piconewton scale forces acting on proteins^{69,70}, cell membranes⁷¹ and DNA⁷²⁻⁷⁴; as well as the effect that antigen spacing have on both single antibodies⁷⁵ as well as whole B-cells during immune system activation⁷⁶. These applications benefit from the ability of DNA nanostructures to position functional groups, most often proteins, at precisely defined spacing and apply them to either *in vitro* or *in vivo* biological systems.

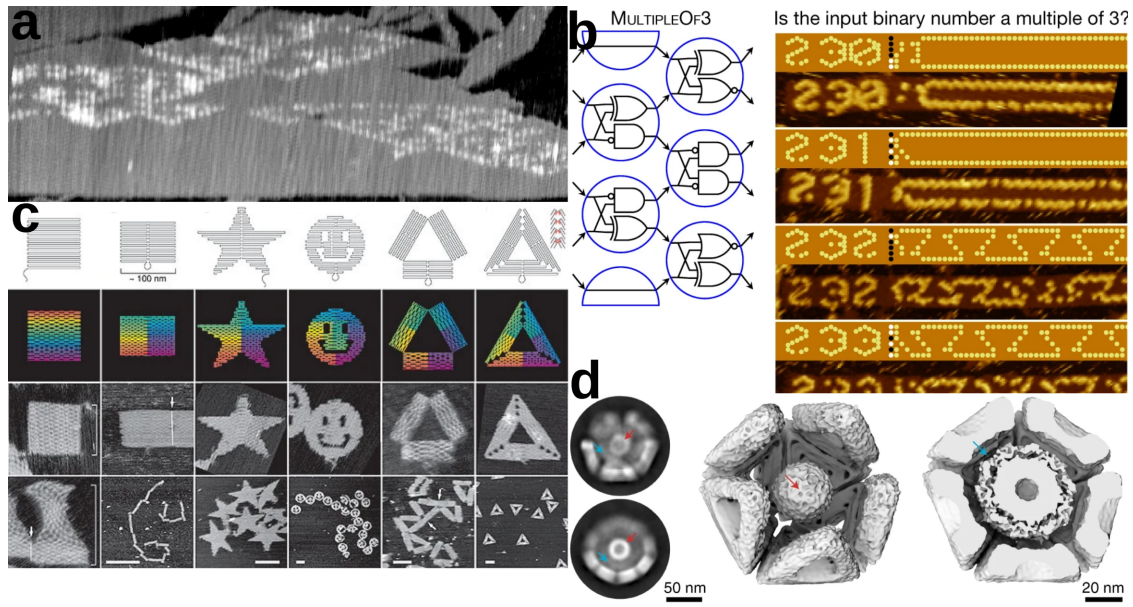


Figure 3. **The evolution of DNA nanotechnology** (a) AFM image of Sierpinski's Triangle algorithmically assembled from DNA by the Winfree lab in 2004. Each DNA tile encoded binding rules which allowed algorithmic growth of the structure. (b) In 2019, the Winfree lab demonstrated a reprogrammable 6-bit DNA computer. They demonstrated numerous algorithms and cellular automata programmed into self-assembling DNA "tapes" which could be seen under AFM. (c) Paul Rothemund's original 2D DNA origami structures from 2006 visualized with AFM. (d) Cryo EM structures of Multi-origami 3D virus-trapping nanocapsules created by the Dietz lab in 2021. They demonstrated multifarious assembly of individual units into multiple different 3D geometries with varying cavity sizes. Images re-used with permission. (a) Used under a creative commons license from⁴⁶. (b) Copyright 2019 Springer Nature⁷⁷. (c) Copyright 2006 Springer Nature⁷⁸. (d) Copyright 2021 Springer Nature⁷⁹.

1.1.4 RNA Nanotechnology

There has also been parallel advancement in RNA nanotechnology. Like DNA, RNA is also made up of a sequence comprised of four chemically distinct bases. Three of them (A, G and C) have the same nucleoside group as corresponding DNA bases, while T is replaced by uracil (U) which is able to form canonical base pairs with both A and G. Besides the substitution of T for U, the only difference between DNA and RNA is the presence of a hydroxyl group at the 2' position of the sugar groups in the backbone. This subtle change at the atomic level leads to a quite dramatic change in polymer structure and a much more diverse and active set of chemical features. At the structural level, rather than forming a B-form helix like DNA, RNA's most common structural motif is the A-form double helix, which has an overall wider structure, longer helical pitch, and a narrow and deep major groove but shallow and wide minor groove. This change is caused by a steric clash

between the 2' OH group and the 5' neighboring base if the ribose sugar is in a 2'-endo conformation, forcing the sugar into the 3'-endo conformation⁸⁰(Figure 5b).

The hydroxyl group also makes RNA substantially more reactive than DNA. This both contributes to its wider functional roles in biology, most notably the peptidyl transferase center of the ribosome (Figure 2b) which is responsible for catalyzing peptide bond formation when assembling proteins, one of the fundamental chemical reactions that make life as we know it possible⁸¹ and self-splicing introns⁸², which are able to self-regulate exon splicing and recombination⁸³. This feature also makes it significantly more difficult to work with in a laboratory setting because interactions between the 2' hydroxyl and the phosphate backbone autocatalyze a hydrolysis reaction which cleaves the RNA strand. This reaction is further accelerated at high pH and in the presence of magnesium ions⁸⁴, which are commonly used in designed nanostructures to screen the charges of the polyanionic phosphate backbone, facilitating the tight packing of helices. There are also the problem of ribonucleases, which many organisms, including humans, secrete into the outside environment as a first line of defense against RNA viruses, necessitating careful handling of RNA solutions.

Design of larger RNA structures is also a challenge. Because of the expense of synthesizing many short RNA strands, creating RNA origami analogous to DNA origami with a single long scaffold and many staples is cost-prohibitive. On the other hand, for small numbers of sequence, RNA can become significantly cheaper than DNA as many copies of an RNA sequence can be prepared from a single DNA template using *in vitro* transcription reactions⁸⁵. For this reason, most engineered RNA structures tend to be made of relatively few strands. It is still possible to create larger RNA structures: by using *in vitro* transcription it is possible to express single RNA strands with thousands of bases which, through creative sequence design, can fold into larger "RNA origami" of similar dimensions to their multistranded DNA counterparts^{86,87}.

Despite these challenges, the promise of RNA nanotechnology is high. Because of RNA's catalytic power and more diverse biological roles, the number of potential applications for designed RNA structures is possibly greater than that of DNA. One of the areas of greatest interest is short interfering RNA (siRNA) therapeutics, which can arbitrarily knock down expression of target genes⁸⁸, including those expressed by viruses or causative of chronic medical conditions. Beyond the siRNA itself, using structured carriers to deliver multiple siRNAs in tandem or using active RNA structures capable of performing computing in order to achieve higher precision delivery are applications where RNA nanotechnology may find a home⁸⁹.

1.1.5 DNA and RNA Nanostructure Design Software

One of the key developments that allowed this growth in complexity was the creation of design software focused on DNA nanotechnology. While it is convenient to think of DNA helices as rods, the helical nature of the molecule itself means that crossovers between neighboring helices can only happen at particular base pair intervals (approximately 10.5 for DNA and 11 for RNA). This, along with the still-developing understanding that the routing of staple and scaffold strands through the structure can have a significant impact on correctly folded product yield⁹⁰, necessitate CAD tools which assist DNA and RNA engineers in designing the global topology of nanostructures.

The first tools for DNA design were command-line based tools which lacked a graphical interface^{91,92}. These were mostly used to produce images for publications and had some rudimentary sequence generation. The first generation of graphical DNA CAD tools came in the 2000s with GIDEON⁹³ and Tiamat⁹⁴ (Figure 4a). These tools provided a 3D editing interface which allowed intuitive creation and connection of DNA (and RNA in the case of Tiamat) duplexes and single-stranded regions. While GIDEON has fallen out of favor, Tiamat is still used by some groups despite having compatibility issues with modern operating systems and has been used to design large and complex DNA origami structures^{54,95,96}.

The next tool released, which remains the most popular, was CaDNAno⁹⁷ (Figure 4b). CaDNAno was the first "lattice-based" design tool, where designs are constrained to parallel helices on either square or hexagonal lattices. This constraint limits the potential design space, but creates a useful abstract language for defining, addressing and routing scaffold and staple strands. Because the helices are parallel, idealized crossover positions can be identified by the software, removing one of the most tedious and prone to user error parts of DNA nanostructure design. This also allows designers to break from the ideal and purposefully create bent and curved structures with confidence⁵¹. CaDNAno revolutionized the DNA design field, however still has some limitations. Primarily, that it constrains DNA designs to parallel lattices. These blocky structures have high structural rigidity and can be assembled into larger structures via either sticky-end cohesion⁹⁸ or blunt-end stacking^{79,99}, but any non-lattice bound components need to be designed by hand and added to the exported sequence list. The fact that all the helices are parallel can also make structures with 3D geometries more difficult to visualize and work with due to large numbers of connections between individual DNA bundles^{98,100}.

Recently, there has been an explosion in the number of design tools as the field continues to focus on larger and more complex structures¹⁰¹. These tools can broadly be grouped into three categories: CaDNAno-like,

automatic routing algorithms, and freeform design tools. In the CaDNAno-like field, the two other programs are Scadnano¹⁰², a browser-based re-implementation of CaDNAno with a more advanced scripting interface to further automate the creation of complex structures and ENS Nano¹⁰³, a design tool which attempts to fix CaDNAno's problem with multi-bundle structures by allowing non-parallel lattices and dynamic relaxation to position the lattices in 3D space.

Automatic scaffold routing tools try to automate the design of nanostructures even further. For these programs, the user begins by defining a polyhedral mesh and then the program attempts to algorithmically fit a DNA origami structure to the mesh. There are three program suites which perform this fitting, first is the BScOR/vHelix package^{104,105}. BScOR contains multiple options for converting closed-surface polygons and flat line drawings into triangulated meshes with mostly single-duplex edges and multiway junctions at the vertexes. The output can then be free-form edited by vHelix, a plugin for the 3D modelling software, Maya. The next group is the Athena suite¹⁰⁶, which includes the fitting algorithms DAEDALUS¹⁰⁷, for 3D wireframes with edges that are an integer number of full turns in length, PERDIX¹⁰⁸ for 2D wireframes with arbitrary edge lengths, TALOS¹⁰⁹, for 3D wireframes with 6-helix bundle edges, and METIS¹¹⁰ for 2D structures with 6-helix bundle edges. These algorithms are focused on optimizing the rigidity of the resulting structure. The final option for routing DNA nanostructures through a mesh is MagicDNA¹¹¹ (Figure 4c), a Matlab program which converts 3D space-filling models into CaDNAno-like bundles of parallel helices. MagicDNA also includes the ability to make structures with sizes that exceed the length of a single scaffold strand and to customize the connections between neighboring helix bundles.

While all the previously-mentioned software packages have focused exclusively on DNA, in the free-form design space the software tools have moved away from Tiamat's vision of fully free-form DNA design towards a focus on tweaking designs produced by other tools and including non-DNA components into the structure. Adenita¹¹² and its in-browser re-implementation, Catana¹¹³ are programs which focus on multiscale representations of DNA and protein structures. They are able to both visualize and edit structures using a CaDNAno-like flat lattice interface, as well as position the lattices in 3D space and perform off-lattice edits. These tools are also able to visualize both proteins and DNA at either a bead-per-nucleotide coarse-grained resolution or fully atomically (though the enormous size of DNA nanostructures often renders fully atomistic visualizations unusably slow). OxView¹¹⁴ (Figure 4d), which will be covered in detail in chapter 2, is a visualization and editing tool designed for the coarse-grained DNA, RNA and protein modeling tool oxDNA/RNA. While its primary function is to visualize simulation results, it is also a fully-featured editor able

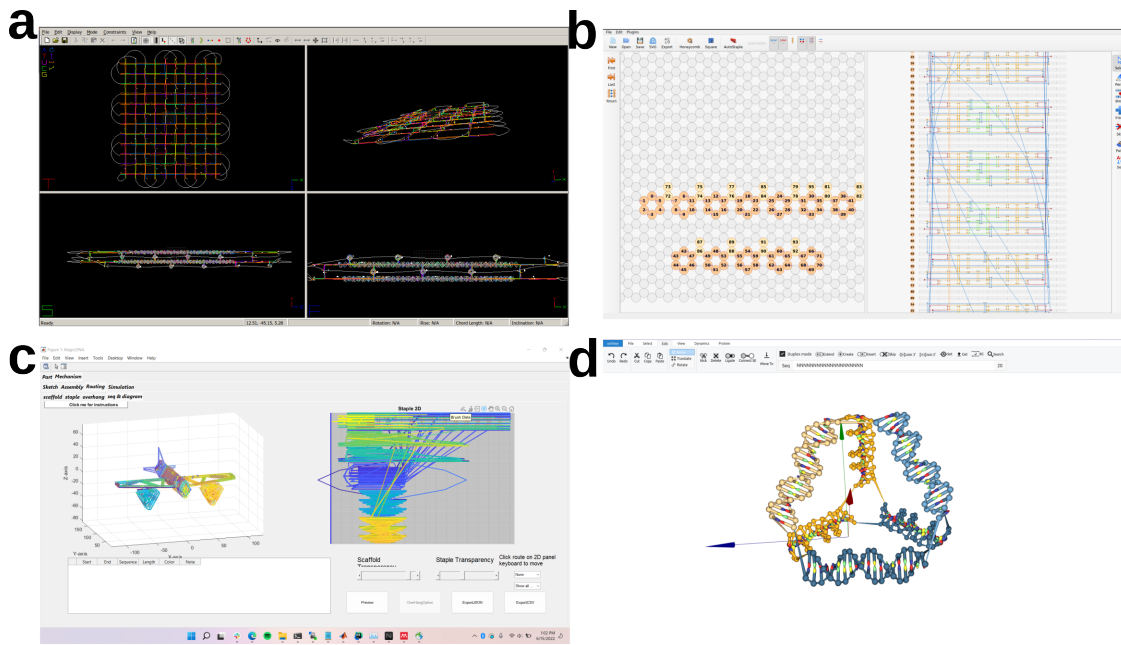


Figure 4. **Examples of design tool interfaces** (a) Tiamat⁹⁴ was used to design a DNA gridiron structure¹¹⁶. (b) CaDNAo⁹⁷ was used to design a DNA octahedron used in crystal lattice assemblies¹¹⁷. (c) MagicDNA was used to design a multi-scaffold design in the shape of an airplane as a demonstration of its multi-scaffold routing algorithm¹¹¹. (d) OxView's¹¹⁴ free-form design tools used to prepare an oxDNA simulation of a DNA tetrahedron⁴¹ created following the protocol described in¹¹⁸. Design files for (a) and (c) retrieved from Nanobase¹¹⁹

to perform free-form modifications of structures and has proven particularly powerful for combining structures designed using different design tools and preparing them for oxDNA simulation.

Currently, no design tool has taken over as the “ultimate DNA design tool” the way SolidWorks is the definitive design tool for macroscale structural materials. Part of this is simply because the field is relatively new and the capabilities and complexity of designs continues to evolve; however, it is also because no tool has been simultaneously fully-featured and intuitive enough to capture the whole field. None of the previously mentioned apps are sufficiently complicated that many of their features couldn't be combined into a slick user interface at some point in the future, perhaps also introducing support for protein design using Rosetta¹¹⁵ as well as for the many small molecules, nanoparticles and inorganic nanomaterials which provide function beyond what is possible for DNA and RNA alone.

1.2 The oxDNA and oxRNA Models

1.2.1 Molecular Models: Applications and Methods

When working at the nanoscale, the interactions of individual monomers and their constituent atoms have a significant impact on the overall structure due to the nanoscopic scale of the construction. However, the ability to visualize nanostructures is limited both by technology and fundamental physics. In ideal conditions, microscopy techniques such as atomic force microscopy¹²⁰ and scanning tunneling electron microscopy¹²¹ can see individual atoms, but these techniques work best for atoms or clusters of atoms deposited on a surface. Nucleic acid structures are far too complex, topologically complicated, and flexible to visualize at the atomic resolution using these techniques. AFM and transmission electron microscopy (TEM) are commonly used to visualize structures, which results in somewhat blurry images where individual helices can be resolved, but not the nucleotides within the helix¹²². The other problem with AFM and TEM is that they both require the imaging target to be on the surface of either a mica or carbon grid, which due to charge interactions between the nanostructures and the surface and the loss of water molecules under vacuum in the case of TEM, cause 3D structures to collapse onto the surface, making these techniques most suitable for 2D structures and extremely rigid 3D structures such as helix bundles. This limitation is somewhat alleviated by cryogenic electron microscopy (cryo EM)¹²³, where samples are frozen in vitreous ice prior to TEM microscopy, which both avoids surface effects and allows for 3D reconstruction of the nucleic acid structure through class-averaging of single particle images, resulting in 3D images with low nanometer to angstrom resolution¹²⁴. Cryo EM requires expensive and specialized equipment and large numbers of images (hundreds of thousands to millions of structures) for high-resolution reconstruction, so is only available to well-equipped labs, however due to its resolution and utility it remains the highest standard of experimental data in nucleic acid nanotechnology.

In silico modelling of structures offers an alternative to microscopy techniques for understanding the dynamics of molecular structures at high resolution. While the models of atomic interaction are never perfect, as the great statistician George Box famously said, "all models are wrong but some are useful"¹²⁵. For modelling structural dynamics of macromolecules, scientists have developed various models ranging from molecular models which dynamically model the interactions between constituent atoms to mathematical models of polymer dynamics which explain bulk properties of polymer chains of various types.

For DNA and RNA nanotechnology, there are two commonly employed types of models, finite-element models and molecular models. The simplest are finite-element models of DNA structure^{126,127} which are able

to predict the overall average structure and flexibility of a given design. These algorithms work by breaking down the structure into a series of quantized elements — base pair steps, crossovers, and single stranded regions which are relaxed from an initial configuration by iteratively adjusting each element’s properties until convergence is achieved. These methods are very fast, taking only a few minutes to predict the overall structure and equilibrium dynamics of a full-sized DNA origami structure. However they are known to struggle with structures with multi-state dynamics and large numbers of single-stranded regions due to the difficult nature of correctly modeling the structural properties of single-stranded nucleic acids.

The second option for computational modeling of structures is either fully-atomistic or coarse-grained molecular modeling where the structural dynamics of the molecule are modeled using physical rules called a “force field”, a mathematical description of the interactions between the individual particles in the system which gives the potential energy for a state. The core theory that molecular simulation works off is the “ergodic hypothesis” which states that given sufficient time, a system will visit all accessible parts of its “phase space” and that all microstates within phase space are found occupied with equal probability. This means that the time average of an “order parameter”, a measurable quantity, will eventually converge to the phase space average. Furthermore, when a studied system is at thermal equilibrium with the outside world, the so-called “canonical ensemble”, many microstates involve reconfigurations in the “heat bath” of the external world, resulting in many microstates with the same energy, E_i . The probability of encountering any particular state is proportional to the exponential of the potential energy of that state divided by the temperature:

$$p(x_i) = \frac{1}{Z} \exp\left(\frac{-E_i}{k_B T}\right) \quad (1.1)$$

Where Z is the “canonical partition function”, — the sum of all possible energy states:

$$Z = \sum_i \exp\left(\frac{-E_i}{k_B T}\right) \quad (1.2)$$

This is the “Boltzmann Distribution”¹²⁸, which explains why low energy states are the most common and higher energy states become more frequent as the temperature rises — there are more reconfigurations of the heat bath which allow states of a particular energy to exist¹²⁹.

The ergodic hypothesis has a suggestive consequence, namely that sampling from allowed states with a probability distribution which follows the Boltzmann distribution or allowing a system to time-evolve using integration of the force field under newtonian mechanics should produce the same result. The former method, so called the “Metropolis Monte-Carlo” (MC, also Markov Chain Monte-Carlo (MCMC)) method¹³⁰, does exactly this. At each step in the simulation, a change to the arrangement of particles is suggested, often within some radius of the previous step’s configuration, the change in energy of the system is then computed from the

force field and if $\Delta E < 0$, the move is accepted automatically while if $\Delta E > 0$, the move is accepted with probability

$$\exp\left(\frac{-\Delta E}{k_B T}\right) \quad (1.3)$$

which is simply the quotient of the probabilities of states i and j with $\Delta E = E_j - E_i$.

The alternative is known as “molecular dynamics” (MD), where the system is allowed to evolve via Newtonian mechanics under the direction of the force field. This is done by calculating the forces on a particle, \mathbf{f}_i , by taking the derivative of the force field, V , with respect to the vector position of particle i , \mathbf{r}_i

$$\mathbf{f}_i = -\frac{\partial V}{\partial \mathbf{r}_i} \quad (1.4)$$

Which can then be used apply an acceleration to particle i by Newton’s Second Law, $\mathbf{f} = m\mathbf{a}$. And allowing the system to evolve linearly for a small amount of time, Δt before recalculating the force field and again applying accelerations to each particle. The most computationally efficient method of computing a molecular dynamics trajectory is the “velocity Verlet scheme”, which computes the position, \mathbf{r}_i , and the velocity, \mathbf{v}_i for each particle at time $t + \Delta t$:

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \mathbf{v}_i \Delta t + \frac{\mathbf{f}_i(t)}{2m} \Delta t^2 \quad (1.5)$$

$$\mathbf{v}_i(t + \Delta t) = \mathbf{v}_i(t) + \frac{\mathbf{f}_i(t) + \mathbf{f}_i(t + \Delta t)}{2m} \Delta t \quad (1.6)$$

Which comes from a Taylor expansion of the positions, resulting in a fast algorithm which is precise up to $O(\Delta t^4)$ ¹³¹.

The major advantage of MD over MC is that MD is trivially parallelizable while MC is not. Since the two methods sample the same underlying distribution, which one to choose is largely a factor of system size. At small system sizes, MC can outperform MD because of the larger distortions to the overall configuration at each simulation step and can be further accelerated for systems of strongly interacting particles (such as DNA and RNA) by using the “virtual move Monte Carlo” algorithm¹³² which proposes moves of groups of particles rather than each particle individually as in traditional MC simulations. However, at large system sizes, due to the advances in highly parallel computing on modern GPUs, MD simulations outperform MC simulations by many orders of magnitude¹³³. MD simulations also have the advantage of, in certain cases, being able to have a correspondence to experimental timescales and kinetic reaction rates due to the continuous nature of the simulations.

Force fields generally depend on the relative positions of the particles and can either be pairwise, depending on each pair of particles, or many-body interactions, depending simultaneously on multiple particle positions (the most common many-body interactions are rotational and torsional potentials which constrain the angles

between groups of particles). Many force fields have been developed over the years with increasing levels of refinement, and accuracy. Of course, more complex and detailed force fields require more computational power and so their use either requires advancement in computational technology or working with smaller numbers of particles. For highly complex quantum force fields, the number of atoms able to be studied remains quite small; for example, one recent study of SARS-CoV-2 M protein inhibition modeled only 75 atoms quantum mechanically while analyzing the mechanism of the interaction¹³⁴. At the classical fully atomistic level, a state of the art RNA force field running on specialized hardware was able to simulate the melting and re-forming of a 14-nucleotide hairpin, which contains 455 atoms¹³⁵. A full-sized DNA origami wherein every nucleotide in the M13 scaffold is paired contains a bit more than 460 000 atoms, 3 orders of magnitude larger than the RNA hairpin.

Due to the enormous numbers of particles in a DNA origami, fully atomistic simulations are often infeasible due to the amount of time required to model that many atoms. To remedy this, many "coarse-grained" simulation methods focused on DNA and RNA have been developed over the years. These models abstract the roughly 30 atoms in each nucleotide into a lower number of coarse-grained particles. This has the effect of decreasing the number of degrees of freedom and therefore increasing the computational efficiency of the model, allowing the same hardware to model larger system sizes and longer timescales. Coarse-graining inherently results in a decrease in accuracy of the model, so it is important to choose a model which accurately reproduces the specific features of interest for the studied molecules. There are two approaches to generating coarse-grained models, bottom-up or top-down (also called empirical parameterization). In bottom-up approaches¹³⁶ to DNA such as IMC DNA¹³⁷, Fan Bonds¹³⁸, or Sugar Model¹³⁹, a model of DNA is built by averaging the contributions observed in fully atomistic models from multiple atoms into a single "super atom". These models have the advantage of being able to represent different parts of the DNA molecule with different levels of resolution. For example, the Sugar Model represents the deoxyribose sugar with more super atoms than the phosphate or nucleoside groups, which allows it to capture the effects of changes in sugar pucker which is responsible for the A-form to B-form helix transition¹³⁹.

Examples of empirically parameterized DNA models include oxDNA¹⁴⁰⁻¹⁴³ and mrDNA¹⁴⁴. These models are developed by designing a super atom with specific interaction sites, and then tuning the interactions between super atoms to reproduce experimentally measured structural properties of DNA. The oxDNA model is by far the most popular of the aforementioned DNA models, having been used in nearly 200 papers since its original publication¹⁴⁵. It was parameterized to capture the helical structure of B-form with accurate geometry, persistence length and melting temperatures. It was later found that though they were not part of the initial

parameterization dataset, accurate kinetics of DNA strand displacement was also an emergent property of the model, solidifying its status as a fast but relatively accurate representation of DNA structure and dynamics¹⁴⁶.

1.2.2 Coarse-Graining DNA, RNA and Proteins

The oxDNA model uses a single rigid, anisotropic bead per nucleotide. Each bead has three interaction sites, the backbone site where it connects to neighboring beads on the same strand, the stacking site, which captures the pi-pi stacking of nucleosides, and the base pairing site which captures the hydrogen bonding responsible for canonical (Franklin-Watson-Crick) base pairing. The backbone is represented as a finite extensible nonlinear elastic (FENE) potential¹⁴⁷, which behaves like a classical Hookian spring near its equilibrium length but becomes more nonlinear as the distance between neighboring particles diverges from the ideal length. Stacking and hydrogen bonding are represented by Morse potentials¹⁴⁸, which is an asymmetric harmonic well which has a distance beyond which it smoothly goes to 0, allowing bonds to both form and break. Excluded volume interactions between non-neighboring nucleotides is in the form of a Lennard-Jones potential¹⁴⁹, which prevents nucleotides from passing through each other and provides some stiffness to single-stranded DNA regions. Electrostatic interactions between backbone sites are given by a Debye-Hückel (DH) potential¹⁵⁰ parameterized to monovalent (Na^+) salt ion concentrations¹⁵¹. It should be noted that most DNA nanostructures are folded at high Mg^{2+} concentrations rather than Na^+ ; however, the effects of Mg^{2+} are highly localized, non-uniform and poorly understood. For structurally constrained DNA junctions, high ($> 0.5 \text{ M}$) Na^+ concentrations give similar structural results to the Mg^{2+} concentrations used in experiments. Finally, a simple quadratic well is applied between diagonally adjacent beads in duplexes as a “cross-stacking” interaction which helps to improve the accuracy of the stacking interactions. In total the oxDNA force field can be written as a sum of pairwise interaction terms between bonded and non-bonded nucleotides:

$$V_{oxDNA/RNA} = \sum_{bonded} (V_{backbone} + V_{stacking} + V_{excludedvolume}) + \sum_{non-bonded} (V_{HB} + V_{crossstacking} + V_{excludedvolume} + V_{coaxialstacking} + V_{DH}) \quad (1.7)$$

OxDNA includes both an average-sequence and sequence-dependent parameterization depending on whether the user would like to study the behavior of a DNA design with or without considering the effect of sequence¹⁴³. In the average-sequence model, all nucleotides are treated as having the same interaction

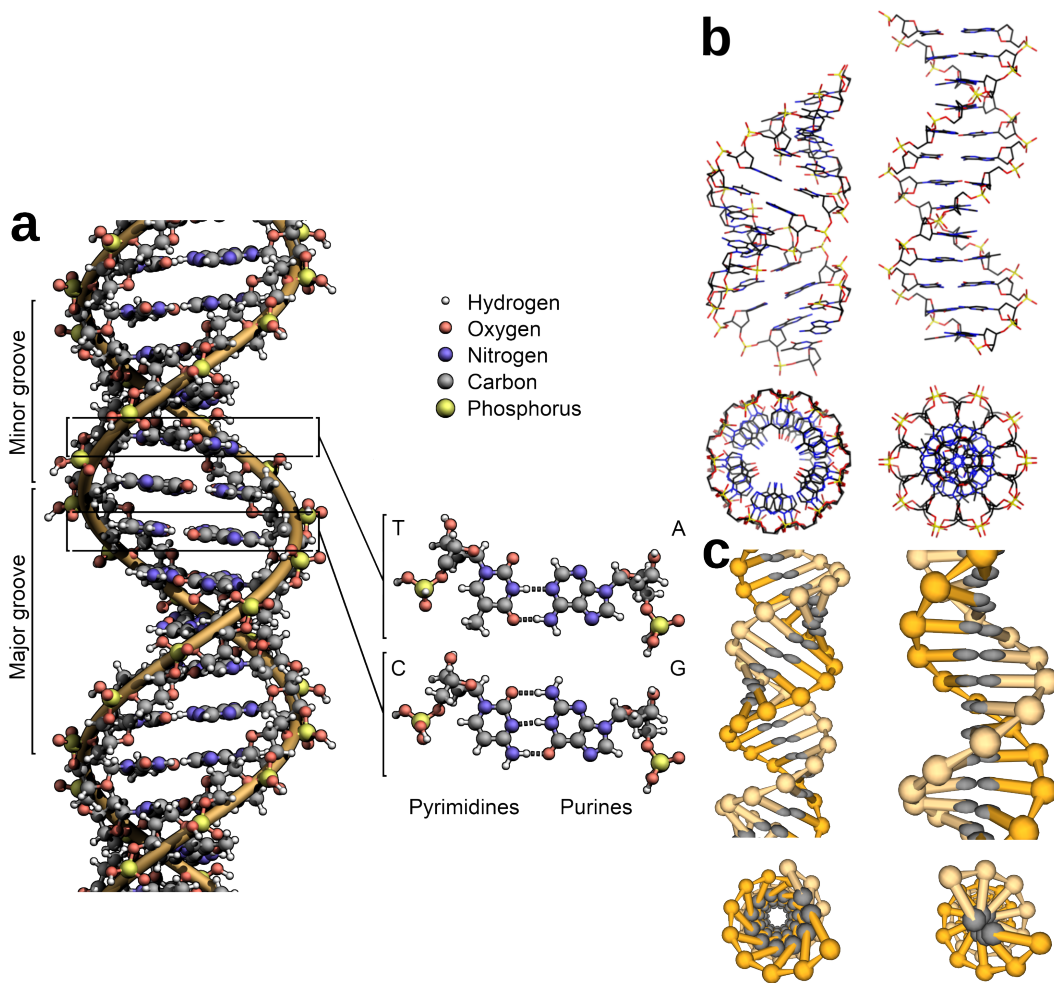


Figure 5. **Structure of DNA and RNA helices.** (a) Fully atomistic detailed view of the DNA B-form double helix. Each monomer along the strands is either A, T, G or C, which have specific base pairing shown in the inset. Pi-pi stacking interactions between the carbon-nitrogen rings of neighboring bases gives rise to the helical structure because these interactions are shorter than the sugar-phosphate covalent backbone (thick orange line). (b) Structure of A-form and B-form helices shown in fully atomistic wireframe models. RNA forms an A-form helix, while DNA forms a B-form helix. Note the difference in helix width and groove geometry which help to differentiate RNA from DNA for interactions with proteins in a biological context. (c) Idealized A-form and B-form helices for the oxRNA and oxDNA models created using oxView. In these coarse-grained models, the 30-40 atoms in each base is replaced with a single, anisotropic bead with empirical interaction potentials which reproduce the geometry and mechanics of RNA and DNA molecules. (a) used under a GPL-3 license. Created by Wikipedia user Zephyris. (b) Used under a CC BY-SA 4.0 license. Created by Wikipedia user Mauroesgurroto.

potential. In the sequence-dependent model, each dinucleotide step (i.e an AT-TA stack of base pairs is different from TA-AT) has thermodynamics given by the Santa-Lucia model¹⁵².

One of the most expensive parts of fully atomistic simulations is modeling solvent molecules. In any biomolecular simulation, the number of water molecules required to fully hydrate the molecule of interest far

outnumbers the number of atoms in the molecule itself. This is one of the largest sources of inefficiencies in molecular simulation, however must be included in order to capture the many roles that water plays in molecular interactions including mediating structural interactions¹⁵³, and entropic effects of structured water¹⁵⁴. In empirically parameterized coarse-grained models, these interactions are generally implicitly included in the force field since they are contributors to the overall structural behavior of the molecule of interest, however this fails to capture the thermodynamic effects of water, namely maintaining the temperature and pressure of the molecule, which diverges from the set point due to the buildup of the error terms from Verlet integration. To remedy this, molecular simulations include “thermostats” which modulate velocities in the simulation to maintain a constant temperature (and similarly “barostats” for pressure). In fully atomistic models, the thermostat generally acts only on the water molecules and the temperature of the molecule is then enforced by collisions with the water molecules. In coarse-grained simulations where water is implicit, the thermostat must act directly on the particles of the studied system.

OxDNA implements multiple thermostat types, however the two of relevance here are Andersen-like^{155,156} and Bussi–Donadio–Parrinello (Berendensen-like)¹⁵⁷ thermostats. In an Andersen thermostat, temperature is maintained by randomly resampling particle velocities from a Boltzmann distribution at a particular temperature after a fixed number of steps. This thermostat has been shown to correctly reproduce the canonical ensemble for a fixed temperature and is the most popular thermostat for production simulation runs. In a Bussi–Donadio–Parrinello thermostat, velocities are randomly rescaled such that the overall distribution of velocities matches the Boltzmann distribution at a given temperature. The probability of rescaling is set by a coupling factor τ , which sets how often, on average, each particle has its velocity rescaled. Bussi–Donadio–Parrinello thermostats with low τ are particularly well-suited to relaxation of structures where overstretched bonds and overlapping particles can result in extremely high forces and thus high particle velocities¹⁵⁸.

The oxDNA simulation engine was later extended to be able to simulate RNA with the oxRNA force field¹⁵⁹. The form of the super atoms and force field are essentially the same as for the oxDNA force field, however the parameters of the force field are modified such that the nucleotides form an A-form helix instead (Figure 5), the persistence length is closer to that of RNA and the thermodynamics of the sequence-dependent model reproduce the Turner Model¹⁶⁰ rather than the Santa–Lucia model. The oxRNA model also includes support for G–U “wobble” base pairs.

Recently, a highly coarse-grained representation of proteins was added to the oxDNA and oxRNA force fields¹⁶¹. The model represents proteins as anisotropic network models (ANMs) where each amino acid is represented by a single super atom at the site of the $C\alpha$ atom. Interactions between amino acids are imple-

mented as harmonic springs with spring constants and equilibrium lengths parameterized on a per-protein basis based on B-factors from crystal structures. The proteins interact with the DNA only through excluded volume (Lennard-Jones) interactions. This model is not meant to be highly accurate for protein dynamics, but to assist researchers designing DNA- and RNA-protein hybrid nanostructures understand the relative sizes of nucleic acid and protein structures and to give a rough idea of the forces that protein fluctuations apply to DNA/RNA structures.

1.2.3 Limitations of the Model

This coarse-grained approximation of DNA and RNA structure does have its limitations, however. Structurally, since only Franklin-Watson-Crick base pairing and stacking interactions are parameterized, hydrogen bonding interactions that use other edges of the nucleotide are excluded. This means that structures containing Hoogsteen- or sugar-edge bonds such as I-motifs, G-quadruplexes, triple helices and any RNA tertiary interactions are not able to be simulated. As previously mentioned, another point of weakness is the parameterization of electrostatic interactions; magnesium interacts with DNA and RNA structures in highly a non-uniform and site-specific manner¹⁶² and clear data on the effects of salt concentration on DNA and RNA stability is only available for sodium buffers¹⁶³. For constrained junctions, high sodium concentrations (0.5 – 1 M) result in very similar overall structures to the standard 5 – 20 mM magnesium concentrations used in experimental setups¹⁶⁴. This does, however have an effect on simulations of individual Holliday junctions, where simulated junctions have a preference for the left-handed conformation while junctions in solution prefer the right-handed conformation. The reasons for this include both the lack of magnesium at the junction point and the simplification of the backbone structure¹⁶⁵.

Finally, coarse-graining inherently decouples timescales of fluctuations in the simulation and oxDNA has an artificially inflated diffusion constant to facilitate simulations of rare interaction events. Based on studies of duplex hybridization, some rates are scaled to be as much as 100x faster than direct unit conversion would suggest¹⁶⁶. This makes determining exact correspondence between simulation and real time infeasible, however relative rates between alternative designs or experimental conditions generally remain accurate.

1.2.4 Choosing the Correct Simulation Methods

Choosing the correct simulation method for a particular task can be a challenge. The simplest option is an equilibrium simulation where the system is simply allowed to evolve under the control of the force field. This type of simulation is useful if the structure is relatively simple and the order parameters which need to be determined are features like the overall flexibility of the structure, the ensemble of distances between a set of nucleotides within the structure, or the equilibrium angle between two duplexes. If, however, the structure is able to undergo some sort of dynamic transition, there is an energy barrier between two possible states, or a full free-energy landscape for an order parameter is needed, "advanced sampling techniques" are needed to more completely sample the order parameter of interest. There are four advanced sampling techniques supported by the most up-to-date version of the oxDNA code.

The simplest of these techniques is "parallel tempering"¹⁶⁷ where multiple simulations are run in parallel at different temperatures. After a pre-determined number of steps, the simulations are paused and the current configurations are exchanged between temperatures with a probability matching the Metropolis MC criteria (see eq. 1.3). In this way, the simulation is able to explore more of phase space as more is accessible at higher temperature. These structures from the high temperature simulations, when swapped into the low temperature condition are often able to seed the lower temperature with energy minima that the low temperature simulations would have taken an unreasonable amount of time to discover due to the intervening energy barriers between them and the starting configuration. When the time comes for analysis, the shuffling is reversed, providing trajectories for each parallel simulated system as it traversed the various temperatures.

Another technique which facilitates sampling across energy barriers is "forward flux sampling"¹⁶⁸, which is used to sample rates of transition between two states, A and B along a particular transition pathway. In forward flux sampling, multiple parallel unbiased simulations are begun from the same initial configuration and each are allowed to run until a pre-determined transition state in the order parameter, λ_0 has been observed a sufficient number of times. The simulations are then stopped and restarted from λ_0 . Simulations that return to A are stopped and restarted at λ_0 until sufficient instances of a second transition state, λ_1 have been observed. This pattern continues until the transition from the final transition, λ_n to B , has been observed the desired number of times. The rate of transition from A to B can then be computed from the product of the transition rates between all of the intervening states.

The final two methods discussed here are "biased" simulation methods where additional forces are added to the simulations to force the system to fully sample a given energy landscape. In "umbrella sampling"¹⁶⁹,

a biasing force (this can be a physical force in MD or a modification to the acceptance probability in MC) is applied to the system to force the simulation to explore only a small area around a particular value of an order parameter (or coordinate in a multidimensional order parameter set). This is repeated in parallel for a number of windows which cover all relevant values of the order parameter. After the simulation is completed, the weighted histogram analysis method (WHAM)¹⁷⁰ is used to unbiased the simulation and compute the free energy profile across the order parameter of interest.

In “metadynamics”¹⁷¹, a simulation is dynamically biased against values of the order parameter(s) of interest that have already been observed. This has the effect of “filling” the free energy profile and can allow efficient sampling of transition states and free energy profiles with multiple minima. This method is particularly useful as it does not require the transition states to be known *a priori* as in umbrella sampling, but instead tends to find the lowest energy barrier between two neighboring states without any pre-knowledge of where that transition is¹⁷².

1.3 Motivation For a Better oxDNA Software Environment

Though the oxDNA/RNA model has been successfully adopted by the computational and theoretical members of the DNA and RNA nanotechnology community, there remain significant barriers to setting up simulations and analyzing the results. This has prevented wider adoption, particularly by experimental groups who would benefit from the analysis and intuition granted by simulations, but do not have the resources to dedicate students full-time to working with the software. The problems with the usability of oxDNA are fourfold:

1. Limited options when converting from DNA/RNA design tools into the model and then relaxing the converted structures to a simulation-ready state.
2. Difficulties compiling the simulation engine and correctly setting up the simulation parameters. This is further exacerbated by the fact that oxDNA is a Unix command line application which itself is a barrier to entry for those who are not already computer savvy, and this is before considering how to correctly set up the necessary compilers and libraries used by the simulation engine.
3. Poor visualization options. Prior to the work outlined in chapter 2, visualization was limited to a converter to UCSF Chimera¹⁷³ which took upwards of half an hour to visualize an origami-sized structure and could not visualize trajectories at all, and Cogli1, a C++ app built to support oxDNA, however it struggled to load large trajectories and had poor documentation.
4. Outdated and confusing analysis tools. The oxDNA package comes with a directory of analysis tools

featuring various functions used in simulation papers published by the maintainers of the code. These tools, however tended to be highly specific and were not of interest beyond the original context. They were also written in Python 2.7 which ended support at the beginning of 2020, necessitating an update.

Item 1 of the preceding list was tackled in ref¹⁷⁴ and chapters 2 and 3 of this dissertation cover software projects intended to remedy these deficiencies. Chapter 4 covers a demonstration of using oxDNA and the tools developed in the previous chapters to study a complex nanodevice constructed from DNA. The software projects include a new, more intuitive and performant visualizer for oxDNA trajectories, a library of analysis tools that cover common use cases in molecular simulation and written in a modular and extensible manner in modern Python, and a web interface for running oxDNA simulations which makes equilibrium sampling simulations more accessible to non-experts.

A key feature of these tools is that they were written in such a way to encourage extensibility and hacking. As software designers, we cannot predict all possible user demands, especially in a field as dynamic as DNA/RNA nanotechnology. As such, these tools should make it easy to re-use their basic features to ask new questions of simulated structures in addition to providing robust core functionality. This design principle also makes them an excellent introduction to programming for simulation analysis that helps to educate and prepare the next generation of scientists. The overarching goal of these projects is to allow the DNA and RNA nanotechnology community to harness the power of simulations to improve their designs and rationalize their results.

1.4 A Co-author's Introduction to Chapter 4

Chapter 4 represents a large collaborative project in which the simulations I performed are only a part of the final product. I chose to include this particular project over other first-author projects^{118,119} of mine because it was the largest and most comprehensive simulation project I undertook during my PhD. It offers a fantastic example of usage for the tools developed in chapter 2 both to build the structures for simulation in the first place and as part of the analysis of the more than 3 TB of simulation trajectories produced as part of the project. The simulations were indispensable in rationalizing experimental results due to the power of oxDNA simulations to explore structures at nucleotide resolution and in 3D. This revealed many features of the nanoengine that were not fully understood from the experimental characterization including the intrinsic twist of the structure, the role of plectoneme formation on RNA polymerase processivity, and the effect non-designed hairpin structures had on both the equilibrium and dynamic behavior of the structure. Furthermore

the ability to tweak the force field to exclude hydrogen bonding interactions from parts of the structure allowed us to explore the effects of the hairpin structures in great detail and propose a solution which could be followed in future research on structures of this type.

My contributions to the text are the sections on molecular dynamics, the remainder of the text was mostly written by Mathias Centola and Michael Famulok. I have included the introduction and the sections on nanoengine design and driver–follower experiments as they are relevant for understanding the results of the molecular dynamics experiments.

1.5 Full List of Publications

For further reading on my contributions to the field of DNA and RNA nanotechnology, please see my complete list of related publications. Entries where my name is in **bold** are first author or co–first author manuscripts.

1. R. Pradeep, J. Procyk, P. Nandi, E. Poppleton, Y. Xu, A. Prasad, F. Zhang, D. Williams, H. Yan, N. Stephanopoulos, P. Chiu, P. Šulc. Coarse-grained simulations for the characterization and optimization of hybrid protein-DNA nanostructures. In revision at *ACS Nano*.
2. J. Bohlin, M. Matthies, **E. Poppleton**, J. Procyk, A. Mallya, H. Yan, P. Šulc. OxView: a system for online design, analysis, and interactive simulation of DNA, RNA, and hybrid protein-nucleic acid nanostructures. (2022) *Nature Protocols*, in press (available online)
3. M. Liu, G. Pedrielli, E. Poppleton, P. Šulc, D. Bertsekas. ExpertRNA: A new framework for RNA structure prediction. (2022). *INFORMS Journal on Computing*, in press (available online)
4. **E. Poppleton**, A. Mallya, S. Dey, J. Joseph, P. Šulc. Nanobase.org: a repository for DNA and RNA nanostructures. (2022) *Nucleic Acids Research*, **50**(D1):D246–D252.
5. D. Kuřák, E. Poppleton, H. Miao, P. Šulc, I. Barišić. Unified Nanotechnology Format: One Way to Store Them All (2022) *Molecules*, **27**(1):63.
6. M. Centola, E. Poppleton, M. Centola, J. Valero, P. Šulc, M. Famulok. A rhythmically pulsing leaf-spring nanoengine that drives a passive follower. (2021) *Biorxiv* Preprint
7. **E. Poppleton**, R. Romero, A. Mallya, L. Rovigatti, P. Šulc. OxDNA.org: A public webserver for coarse-grained simulations of DNA and RNA nanostructures. (2021) *Nucleic Acids Research*, **49**(W1): W491–W498.

8. J. Procyk, E. Poppleton, P. Šulc. Coarse-grained nucleic acid-protein model for hybrid nanotechnology. (2021) *Soft Matter*, **17**(13):3586-3593.
9. G. Yao, F. Zhang, F. Wang, T. Peng, H. Liu, E. Poppleton, P. Šulc, S. Jiang, L. Liu, C. Gong, X. Jing, X. Liu, L. Wang, Y. Liu, C. Fan, H. Yan. Meta-DNA Structures. (2020) *Nature Chemistry*, **12**(11):1067-1075.
10. J. P. K. Doye et al. The oxDNA coarse-grained model as a tool to simulate DNA origami. (2020), *ArXiv* preprint
11. **E. Poppleton**, J. Bohlin, M. Matthies, S. Sharma, F. Zhang, P. Šulc. Design, optimization and analysis of large DNA and RNA nanostructures through interactive visualization, editing, and molecular simulation. (2020) *Nucleic Acids Research*, **48**(12):e72.
12. A. Suma, E. Poppleton, M. Matthies, P. Šulc, F. Romano, A. A. Louis, J. P. K. Doye, C. Micheletti, L. Rovigatti. Tacoxdna: a user-friendly web server for simulations of complex DNA structures, from single strands to origami. (2019) *Journal of Computational Chemistry*, **40**(29):2586-2595.
13. M. Matthies, N. P. Agarwal, E. Poppleton, F. M. Joshi, P. Šulc, T. L. Schmidt. Triangulated wireframe structures assembled using single-stranded DNA tiles. (2019) *ACS Nano*, **13**(2):1839-1848.

Chapter 2

SOFTWARE TOOLS FOR PREPARATION, VISUALIZATION, AND ANALYSIS OF OXDNA/RNA SIMULATIONS

This chapter was published in **E. Poppleton**, J. Bohlin, M. Matthies, S. Sharma, F. Zhang, P. Šulc, **Design, optimization and analysis of large DNA and RNA nanostructures through interactive visualization, editing and molecular simulation** *Nucleic Acids Research* 48(12), e72 (2020)

Abstract

This work seeks to remedy two deficiencies in the current nucleic acid nanotechnology software environment: the lack of both a fast and user-friendly visualization tool and a standard for structural analyses of simulated systems. We introduce here oxView, a web browser-based visualizer that can load structures with over 1 million nucleotides, create videos from simulation trajectories, and allow users to perform basic edits to DNA and RNA designs. We additionally introduce open-source software tools for extracting common structural parameters to characterize large DNA/RNA nanostructures simulated using the coarse-grained modeling tool, oxDNA, which has grown in popularity in recent years and is frequently used to prototype new nucleic acid nanostructural designs, model biophysics of DNA/RNA processes, and rationalize experimental results. The newly introduced software tools facilitate the computational characterization of DNA/RNA designs by providing multiple analysis scripts, including mean structures and structure flexibility characterization, hydrogen bond fraying, and interduplex angles. The output of these tools can be loaded into oxView, allowing users to interact with the simulated structure in a 3D graphical environment and modify the structures to achieve the required properties. We demonstrate these newly developed tools by applying them to design and analysis of a range of DNA/RNA nanostructures.

2.1 Introduction

The field of nucleic acid nanotechnology³⁵ uses DNA and RNA as building blocks to construct nanoscale structures and devices. Using the high programmability of pairing combinations between oligonucleotides, it is possible to construct 2D and 3D nanostructures up to several thousand nucleotides. Over the past three decades, designs of increasing complexity have been proposed, such as DNA/RNA tiles and arrays¹⁷⁵, DNA multi-bundle origamis⁹⁹, wireframe nanostructures^{107,176} single-stranded tile (SST) nanostructures¹⁷⁷, single-stranded DNA (ssDNA) and RNA (ssRNA) origami structures⁸⁶, and larger multi-origami tile assemblies⁵³. The nanostructures have promising applications ranging from photonic devices¹⁷⁸ to drug delivery⁵⁹.

There are many available nucleic acid nanotechnology design tools, including CaDNAno⁹⁷, Tiamat⁹⁴, vHelix^{104,105}, Adenita¹¹², MagicDNA¹¹¹ and the CAD converters DAEDALUS¹⁰⁷ and PERDIX¹⁰⁸. CaDNAno is frequently used to design very large structures on either a square or hexagonal lattice, which requires components be made of parallel helices. Tiamat is an intuitive lattice-free design tool that supports both DNA and RNA. MagicDNA is a Matlab-based tool that specializes in the design of large 3D structural components on a 3D cubic lattice using CaDNAno-like parallel DNA bundles as the base unit of each edge. VHelix and Adenita are DNA design plugins for the commercial design platforms Maya and SAMSON. VHelix facilitates conversion of polyhedral meshes to DNA sequences, with further free-form editing available in Maya. Adenita combines the functionality of CAD converters with free-form design, allowing users to load structures from a variety of sources with additional editing tools available in the SAMSON interface. DAEDALUS and PERDIX are software that facilitate conversion of meshes designed in CAD software into DNA representations. Currently, the nanotechnology field lacks a universal method for assembling structures made in different design tools, especially if small changes need to be made. Continued development of tools is thus necessary to integrate the previous efforts and enable design of more complex DNA and RNA nanostructures. Additionally, with the exception of Tiamat, all available tools focus only on DNA nanostructure designs.

Molecular simulations have proved indispensable in the field of nucleic acid nanotechnology, providing detailed information about bulk structural characteristics^{179,180}, folding pathway kinetics^{181,182}, conformational space and kinetics of complex nanostructures^{165,183,184}, and active devices such as DNA walkers^{183,185}. Due to the size of the designed nanostructures and the laboratory timescales involved, traditional fully atomistic simulation methods are often infeasible for nucleic acid nanotechnology applications. To remedy this, several coarse-grained models have been developed^{126,140,141,143,144,159,186-188}, each of which with a unique focus on a specific part of the DNA nanostructural design and characterization pipeline. In particular, the oxDNA/oxRNA models have grown in popularity in recent years and have been used for studying DNA/RNA nanostructures and devices^{165,186,189-191} as well as RNA/DNA biophysics^{141,192,193}. The models represent each nucleotide as a single rigid body, where the interactions between nucleotides are empirically parameterized to reproduce basic structural, mechanical and thermodynamic properties of DNA and RNA (Figure 6).

However, the standalone simulation package only provides simulation trajectory with recorded 3D positions of all nucleotides in the simulation. Users usually have to develop in-house evaluation tools that post-process the simulation trajectory to extract desired properties of the studied nanostructures.

In this paper, we present two open-source tools to fill these unmet needs in the field of DNA/RNA nan-

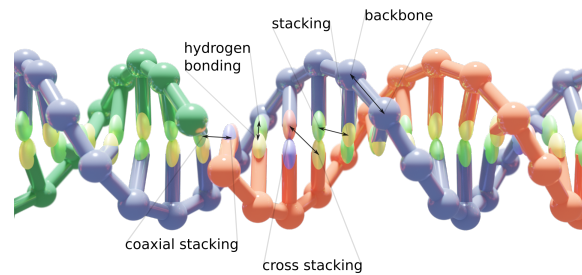


Figure 6. **The oxDNA model.** A DNA duplex as modeled in oxDNA with labels corresponding to the potentials defining the force field. OxDNA is a coarse-grained model with each nucleotide represented as a rigid body with specific interaction sites that approximate the geometry and interactions of the 20+ atoms that make up each nucleotide. The coarse-grained force field is parameterized to reconstruct the structural and dynamic properties of both single- and double-stranded DNA and RNA.

otechnology and illustrate their use for design and optimization of DNA and RNA nanostructures. The first tool we introduce here is oxView, a browser-based visualization and editing platform for DNA and RNA structural design and analysis of nanostructures simulated in oxDNA/oxRNA. The tool is able to accommodate nanostructures containing over a million nucleotides, which is beyond the reach of most other visualization tools. It allows the user to load multiple large nanostructures simultaneously and edit them by addition or deletion of individual nucleotides or entire regions, providing a way to create new, more complex designs from smaller, individually designed subunits, even from different design tools. All of the previously mentioned design tools can be converted to the oxDNA format using either built-in tools (Adenita, MagicDNA, vHelix), the TacoxDNA webserver¹⁷⁴ (CaDNAo, Tiamat, vHelix), or by converting first to PDB using built-in tools and then to oxDNA using TacoxDNA (DAEDALUS, PERDIX). The visualization tool is integrated with oxDNA/oxRNA simulations and loads long simulation trajectories quickly (including files which are tens of gigabytes in size) for interactive analysis and video export of nanostructure dynamics. It can also load data overlays from the analysis scripts introduced in this paper, allowing users to interactively explore features such as hydrogen bond occupancy and structure flexibility and then use this information to iteratively redesign nanostructures based on simulation feedback using oxView. Finally, oxView implements rigid-body dynamics code so that individual parts of the structures can be selected and interactively rearranged. The structure will then be relaxed on-the-fly using rigid-body dynamics to a conformation which can be used as an initial structure in simulations.

The second tool introduced here is a set of standardized structure-agnostic geometry analysis scripts for oxDNA/RNA which cover a number of common molecular simulation use cases. Many groups that work with oxDNA/RNA have developed their own analysis tools in-house, resulting in many duplicate functionalities

and scripts that are limited to single experiments. To facilitate the simulation-guided design of DNA/RNA nanostructures and lower the barrier of entry into the simulation field, we have developed a toolkit that is easy to use, generically applicable to numerous studied systems, and extensible. The tool set includes the following: (i) calculation of mean structure and root-mean squared fluctuations to quantify structure flexibility; (ii) hydrogen-bond occupancy to quantify fraying and bond breaking during the simulation; (iii) angle and distance measurements between respective duplex regions in a nanostructure; (iv) a covariance-matrix based principle component analysis tool for identification of nanostructure motion modes and (v) unsupervised clustering of sampled configurations based on structural order parameters or global difference metrics.

We demonstrate the versatility of the analysis tools and visualization platform functionality by analyzing simulations of previously published structure and a few novel designs. In particular, we study two RNA tiles, a Holliday junction, the tethered multi-fluorophore structure, two wireframe DNA origamis, and a single-stranded RNA origami nanostructure. We make no custom modifications to the analysis tools for each of the designs to demonstrate their versatility and general utility for distinct nanostructures. The visualization and analysis software developed in this work is freely available under a public license.

2.2 Materials and Methods

2.2.1 System and Software Requirements

The analysis tools were written and tested using the following dependencies: Python 3.7 (minimum version 3.6), NumPy 1.16¹⁹⁴, Matplotlib 3.0.3 (minimum version 3.0)¹⁹⁵, BioPython 1.73¹⁹⁶, SciKitLearn 0.21.2¹⁹⁷ Pathos 0.2.3¹⁹⁸, oxDNA 6985 (minimum version June 2019)^{143,185,186}.

OxView will run as-is on any modern web browser with WebGL support; though, we note that Google Chrome performs best at very large structure sizes. To make modifications to the code, the following dependencies are required: JavaScript ES6, Typescript 2.9.0

2.2.2 Simulation Details

The oxDNA simulations of systems that were used in this work have been carried out using the standard molecular dynamics and Monte Carlo approaches. The simulation parameters and file formats produced by the simulations are described in appendix A.

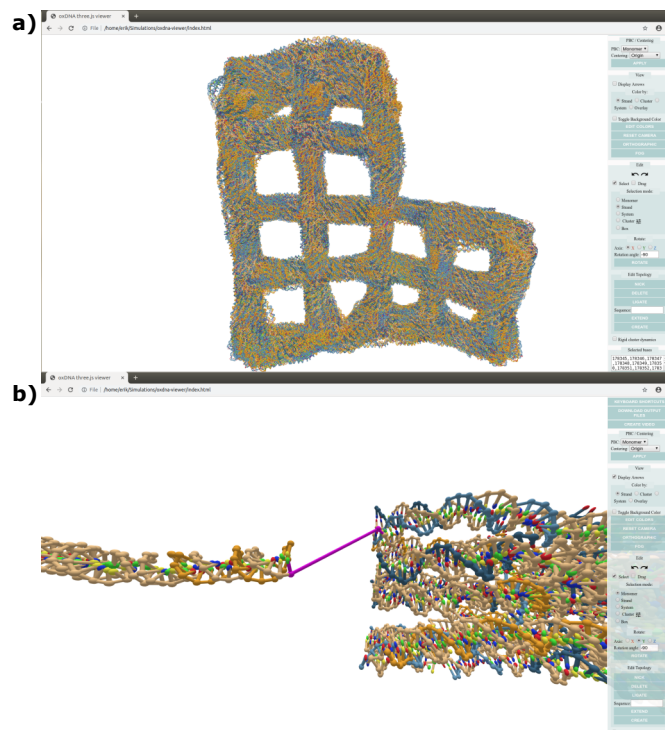


Figure 7. **Screenshots from usage of oxView.** **a)** 100 configurations from an oxDNA simulation of design 24 from¹⁰⁸ merged into a single file and loaded into oxView; illustrating the ability to smoothly visualize over 10^6 nucleotides. The origami design has 11 382 nucleotides, resulting in a combined file containing 1 138 200 nucleotides, which renders as 5 691 000 individual objects in the scene. **b)** Using oxView to assemble a simulation of the tethered multifluorophore (TMF) structure used in¹⁹⁹. Each of the subunits is a separate CaDNAno file converted into oxDNA format using¹⁷⁴. The two subunits and the algorithmically generated tether had to be ligated prior to simulation.

2.3 Results

2.3.1 OxView - Web Browser Visualization, Analysis and Editing of Nanostructures

We introduce oxView, a JavaScript app built on the Three.js visualization library to provide fast, user-friendly, and flexible visualization capabilities with low technical overhead (Figure 7). OxView uses hardware instancing to offload most calculation of object geometry to the computer's GPU, allowing it to smoothly visualize structures containing millions of nucleotides (Figure 7a). Standard Three.js scenes encounter a bottleneck in the rate of CPU draw calls with only a few thousand objects on the screen. By using instanced materials and custom properties written into the WebGL shaders, oxView bundles many objects with similar geometries into a single draw call that calculates edges and vertices in the compiled shader code.

Loading a simulation is as simple as dragging and dropping a trajectory/topology file pair onto a browser window with the app running. Simulation trajectory files can be stepped through using onscreen buttons or the keyboard, and the trajectory movie can also be downloaded as a video file. Available formats are .webm, .gif, and .jpg/.png image archives.

In addition to visualization, oxView also has basic editing capabilities (Figure 7b) and Supplementary video 2). Particles can be selected individually, or whole strands and systems can be selected as a whole. Box selection, range selection (shift+click) and cluster selection are also available. Clustering can be done automatically using a Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm²⁰⁰ or can be assigned manually from other selection methods. Briefly, DBSCAN compares distances between points and classifies groups of points meeting a specified minimum size and within a specified mutual minimum distance as members of the same cluster. It also characterizes points as central or peripheral, with central points having at least the minimum number of neighbors in the cluster and peripheral points being within the cutoff distance of one or more, but fewer than the specified number of central points. When manually selecting nucleotides, holding the control key while making a selection will combine the new and previous selections. Selected particles can be translated and rotated, and the topology can be edited via strand extension and creation, nicking, deletion, and ligation. Edits can be undone and redone using the standard ctrl-z/ctrl-shift-z keyboard shortcuts. Strand extensions will attempt to approximate either an A-form or B-form helix depending on the parent nucleotide's identity: RNA or DNA. The final edited version can be downloaded as an oxDNA file pair for further simulation or as a CSV sequence list for experimental validation.

We envision this tool being used to prototype DNA/RNA nanostructural designs in an iterative process before realization in the lab. The structure can be simulated for a short time, analyzed for defects, and then iteratively modified in the viewer and returned to simulation to verify success. This tool is also useful as a neutral ground between structures designed in other editing tools, allowing researchers to merge together structures from many sources to realize a complex vision.

OxView also allows the creation of mutual trap external force files for oxDNA/RNA. These files define artificial pairwise spring potentials between nucleotides that can be loaded in an oxDNA simulation and be very helpful when simulating the relaxation of a complex structure, assembled from multiple components, or when relaxing a structure imported from the CaDNAno format.

2.3.1.1 Implementation Details

The underlying architecture of oxView has two parallel data streams. The first mirrors the physical arrangement of nucleic acid monomers into strands, with each configuration/topology pair representing a system. This data structure contains the topological information relating to particle identities, connectivity, and relation to the system. Monomers, strands and systems all inherit from the Three.js Group object and are related through an inheritance hierarchy, which allows interaction with structural units as a group. Additionally, each system contains a set of data arrays that define the positions, orientations, sizes, and colors of every particle. These arrays are passed into a custom implementation of the WebGL Lambert shader, where they are compiled on the GPU and drawn as a single object. This scheme allows loading of over 1 million nucleotides into a single scene (Figure 7a and Supplementary video 1).

Selection is handled through a GPU-picker, which avoids the need for computationally-expensive raycaster intersection calculation. Briefly, each nucleotide has a mesh with a color corresponding to its global ID at the same position as its backbone site which is rendered in an invisible scene. The color of this mesh can quickly be determined via the $x - y$ coordinates of the mouse on the screen. When the color is converted from the hexadecimal color to the corresponding decimal value, it returns the ID of the nucleotide under the mouse pointer. As the arrays passed to the shader are of constant-size, new nucleotides added to the scene after initialization, are placed in a temporary system object with its own instancing arrays.

2.3.1.2 Data Overlays in oxView

Many of the simulation analysis scripts introduced in this work output overlay files that can be viewed in oxView. This allows interactive visualization of different properties (such as flexibility, discussed in Figure 9) of respective parts of the structure obtained from simulations. These are JSON-format files that define the name of the overlay and the data. There are three types of overlays recognized by oxView. The most frequently used is the color overlay. These files contain one value per particle. When dragged and dropped into oxView, along with the corresponding configuration/topology pair, the color overlay file will create a superimposed colormap on the structure based on the value associated with each particle. All 256-value colormaps from Matplotlib¹⁹⁵ are available in addition to the default Three.js colormaps. The displayed colormap can be altered via a simple API implemented in the browser console. In addition to per-nucleotide coloring, oxView can also read two JSON formats corresponding to arrows drawn on the scene. The first is a three-component vector for each

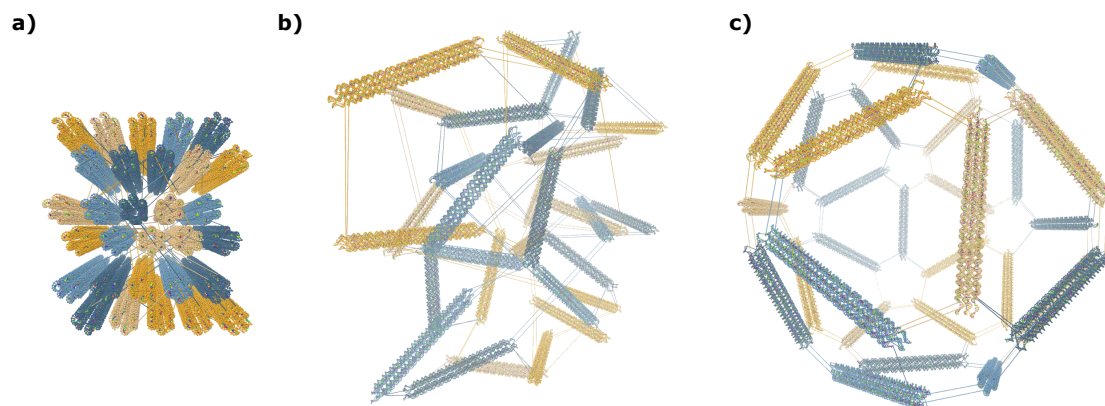


Figure 8. **Rigid-body dynamics of clusters.** Snapshots from the automatic rigid-body relaxation of an icosahedron, starting with the configuration converted from caDNAano **a)**, through the intermediate **b)** where the dynamics are applied, and **c)** the final resulting relaxed state.

nucleotide, which is produced by the principal component analysis script and draws a vector, emanating from each particle, using the magnitude and orientation defined in the overlay file. The second format, which can contain any number of vectors, takes pairs of three-component vectors and draws arrows of the corresponding position and orientation on the scene.

2.3.1.3 Relaxing Structures Using Rigid Body Dynamics

There has been a recent push to develop software that converts structures designed in the various design tools to simulation formats¹⁷⁴. Due to the lattice-based drawing platform with parallel helices used by CaDNAano, exported structures can be very difficult to relax to a physically reasonable state in oxDNA. Initial configurations imported from CaDNAano (shown in Figure 8a) will generally be planar with highly stretched bonds between individual structural units. Thus, without 3D information on how to reorient the helices, neither MC nor MD simulations are able to find the relaxed arrangement. This can also lead to topological impossibilities, where structures are knotted in a nonphysical manner. Additionally, starting simulations from a state with very stretched bonds can result in numerical instabilities that crash the simulation. For origami structures consisting of multiple origami blocks, connected by initially stretched backbone bonds, rigid-body manipulation has previously been used to arrange the converted oxDNA structure into a more realistic initial configuration²⁰¹. The translation and rotation tools in oxView allow users to select and rearrange blocks of nucleotides as rigid bodies. Furthermore, oxView also includes a rigid-body dynamics (RBD)²⁰² mode, that

automatically transforms groups of nucleotides based on a simple force field. It is also possible to drag and rotate groups during RBD, allowing the user to nudge the design into the desired topology. Groups can either be created manually via the selection interface or through the implemented DBSCAN algorithm²⁰⁰ that automatically identifies and categorizes spatially separated groups of particles. The latter option works particularly well with designs developed in CaDNAno.

Each group is represented as a rigid body with a position and an orientation. The groups are held together with spring forces at each shared backbone bond, with a magnitude of

$$f_{spr} = c_{spr}(l - l_r),$$

where c_{spr} is a spring constant, l is the current bond length and l_r is the constant relaxed bond length. To avoid overlaps, a simple linear repulsive force, of magnitude

$$f_{rep} = \max\left(c_{rep}\left(1 - \frac{d}{r_a + r_b}\right), 0\right)$$

is added between the center of each group, where c_{rep} is a repulsion constant, d is the distance between the two centers of mass, and $r_a + r_b$ is the sum of the group radii (the greatest distance they can be while still overlapping). An example of the dynamics in action can be seen in Figure 8 and Supplementary Video 3, where each side of a DNA icosahedron⁵⁰ is automatically arranged into the intended shape.

2.3.2 General-Purpose Analysis Tools

Popular molecular simulation tools programs, such as GROMACS²⁰³, not only perform molecular simulations, but also include analysis tools for common use-cases. The access to reliable and maintained tools, as part of the distribution, allows for standardization between many researchers using the core tool, as well as simplifying the learning curve for new researchers working with the tool. At this time, although there are over a hundred publications using oxDNA/RNA, no standardized set of tools for structural analysis has emerged. We present here a set of tools covering many common structure analyses: mean structure, root mean squared fluctuations (RMSF), hydrogen bond occupancy, interaction energy, interduplex angles, contact mapping, the distance between nucleotides, and principal component analysis of structure motion. These are primarily written in Python, with some portions embedded in the oxDNA C++ code for enhanced speed. Moreover, we provide additional utilities including a parallelization scheme for analyses, trajectory alignment, and unsupervised clustering based on data outputs.

2.3.2.1 Mean Structure Determination and RMSFs

This package includes two methods for determining the mean structure. One utilizes the Biopython¹⁹⁶ singular value decomposition (SVD)-based structure superimposer. This is a popular method²⁰⁴ that finds a translation and rotation to superimpose two distinct conformations on top of each other to minimize the the root mean square distance between their components. Either a user-defined or random configuration in the trajectory is selected as the reference structure. In the example structures displayed here, this choice was found to have little impact on the final outcome. Each configuration is then superimposed onto the reference, and the average position of each nucleotide is calculated by taking the mean of each particle's coordinates in the aligned reference frame. The alignment can also be performed on a subset of particles in the structure. These are assigned from a space-separated index file that can be produced by clicking the 'Download Selected Base List' button in oxView. Sometimes, a mean structure is undesirable because they are frequently not physically possible state. To obtain a physically reasonable, but representative structure, this package also includes a centroid-finding script which finds the structure in a trajectory that has the lowest total RMSF to the a provided reference (such as a mean structure). To find the per-particle RMSF, a second script uses the mean structure produced by the first script as the reference configuration for alignment. The squares of the distances between the alignment and the mean structure for each nucleotide are then summed and divided by the total number of configurations. The square root is then taken to find the RMSF per particle in nanometers. The final output from this script is a .json format color overlay that can be loaded into oxView.

As noted in¹⁶⁵, averaging methods that use full structure alignment work very well for rigid structures; However, there are some caveats. Large planar structures frequently appear to have the smallest RMSF in a ring midway between the center and the edge (Figure 9b). This does not correspond to lower flexibility, but instead reveals an artifact of the single-value decomposition. If a structure can bend in two possible directions, the stationary point in the oscillation will appear to have very low flexibility. Highly flexible regions tend to collapse towards a center line, which is particularly problematic for rigid structures connected by a flexible linker, exemplified by the interrupted duplex shown in Figure 5A. When the average structure is computed for this design, the entire structure collapses into a linear blob that does not have any resemblance to any of the individual configurations. This is because the average position for these flexible particles is drawn towards the center. For such structures, another mean structure calculation based on interparticle distance is employed.

The second option for mean structure determination uses a common machine learning technique, multidimensional

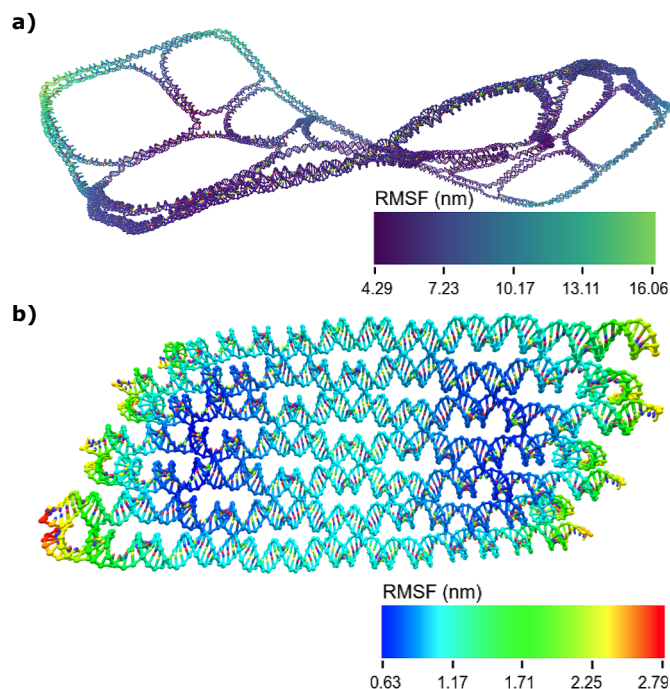


Figure 9. **Mean structures and RMSF.** **a)** The mean and deviations scripts were used to compute the mean structure and RMSFs of design 19 from¹⁰⁸. In the initial report of these designs, they were characterized by AFM, showing complete, flat structures. In the simulations here, the structures were stable; however, the mean structure shows a significant right-handed global twist. **b)** To demonstrate the patterns that appear in RMSF calculations, this is the mean structure of a single-stranded RNA origami⁸⁶ with the RMSF shown using a colormap with high spectral contrast. The center of the origami appears to have an RMSF twice as high as the surrounding regions. This is simply an artifact of the alignment and not an accurate characterization of particle motion.

mensional scaling (MDS)²⁰⁵, to reconstruct a mean structure from local contact maps. MDS is one of a class of algorithms known collectively as manifold learning, which are traditionally used to perform dimensionality reduction in high-dimensional datasets. MDS takes a set of pairwise distances between points in an arbitrary number of dimensions, as an input. The algorithm then uses eigenvalue decomposition to find distances d_{ij} in the embedded space that minimize

$$f(\delta, d) = \sum_{i,j=1}^N (\langle \delta_{i,j} \rangle - d_{ij})^2,$$

where N is the number of data points, $\langle \delta_{i,j} \rangle$ is the mean distance between centers of mass of nucleotides i and j (averaged over the whole simulated trajectory) and $d_{i,j}$ is their embedded distance¹⁹⁷. In the implementation presented here, pairs of nucleotides, where average distance $\langle \delta_{i,j} \rangle$ is longer than the cutoff of $r_{cut} = 2.07nm$ (approximately the interhelix gap in an origami), are not considered in the embedding. The MDS-based mean structure calculation uses the MDS algorithm²⁰⁶, implemented in the Python machine learning toolkit,

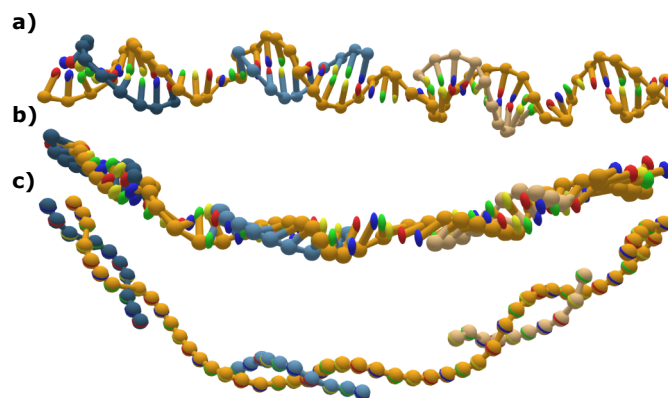


Figure 10. **Improving mean structures of flexible designs.** **a)** The initial configuration of a 50-nucleotide duplex interrupted with 5 nucleotide gaps, created using the editing tools in oxView. Each individual configuration encountered during simulation displayed helical geometry. **b)** The mean structure computed using SVD of the whole simulation. Because of the high backbone and rotational flexibility of this structure, it collapses into a linear shape that has little correspondence to the double helix geometry that is maintained throughout the simulation, **c)** The mean structure computed using MDS. In this case, since only local contacts are used to construct the mean structure, the helical geometry is maintained. MDS comes at the cost of losing nucleotide orientation information, however. Thus, the visualization only shows the center of mass for each nucleotide.

SciKit-Learn¹⁹⁷, to reconstruct these local distances into a three-dimensional embedded representation. This method loses orientation data, and thus, nucleotides are simply visualized as spheres at their centers of mass (Figure 10). Once a mean structure (in the embedded space) is calculated, the script then calculates the mean deviation in distance between each particle and its nearest neighbors and outputs an oxView color overlay file to quantify the flexibility.

We used the SVD-based mean structure script to study flexibility and curvature in large wireframe origami structures¹⁰⁸. In the original research, these structures were visualized using atomic force microscopy (AFM), which tends to overestimate the flatness of structures due to electrostatic interactions between the mica surface and the DNA origami¹⁷⁶. Though the wireframes appear flat in the published AFM results, our simulations suggest that in solution they would be more crumpled or have some degree of global helical twist. Particularly striking is the helical shape of the mean structure of design number 19 from¹⁰⁸ (shown in Figure 9a and Supplementary video 4). OxDNA was parameterized to correctly reproduce the global twist of large 3D DNA structures^{51,143}, suggesting that this twist is likely significant while in solution. We note, however, that the global twist of 2D DNA nanostructures in the bulk remains a topic of active research²⁰⁷, and more experimental data is needed to establish a better comparison of oxDNA parametrization with experimentally determined structures. Mean structures are also the best method to compare simulation results to cryo-EM maps. Both

produce an averaged structure over thousands of individual snapshots. Thus, converting mean structures to PDB format using existing conversion tools¹⁷⁴ for use with cryo map fitting software, such as can be found in Chimera¹⁷³, is a method to correlate simulations and experimental data.

Because of the limitations of SVD-based mean structure calculation, the MDS approach was also used to determine the mean structure and deviations. Unfortunately, because average distance data is noisy and does not precisely map to a single configuration, this method does not work for structures larger than a few thousand particles. In all tests of the algorithm at origami scales, every particle was placed at the origin, a trivial solution that is a known issue of manifold learning methods. However, at smaller scales, this method provides a reasonable mean structure, that respects the geometry of the double helix, and a measure of deviation that reveals areas of flexibility without global artifacts due to fitting (Figure 11).

2.3.2.2 Geometric Parameters: Interduplex Angles and Distances

The simplest structural unit of nanotechnology structures is the duplex — antiparallel strands of sequentially bonded nucleotides. We have implemented a script that automatically determines the duplexes present in each configuration within a trajectory and fits a vector through the axis of the duplex. This is trivial for DNA, where the center points of each base pair lie roughly co-linear and the axis can be defined by a linear regression through the points in the center of the duplex. For RNA, the A-form helix is slightly more difficult to characterize. The duplex is defined by the normal vector to an average plane fit through the displacements along the backbones as described in^{159,208}. This script creates a text file that contains information about all duplexes found at each step. This can be visualized using a separate script, which uses the ID of nucleotides at the edge of the duplex, found using oxView's selection feature. This method can compare angles either within or between structures.

Determining the angle between two duplexes can be useful in assessing design outcomes as well as quantifying twist within nanostructures. The output from the angle script is a list of all duplexes found in each configuration of the trajectory. This output can then be fed into the partnered visualization script along with the starting nucleotide IDs of the duplex. The output will be the median, mean and standard deviation of the angle between the two duplexes, as well as the fraction of analyzed configurations in which that pair of duplexes are both present. This number is an indication of both how stable the structure is and whether or not the chosen duplex is representative of the entire trajectory. The script will also provide a histogram and/or

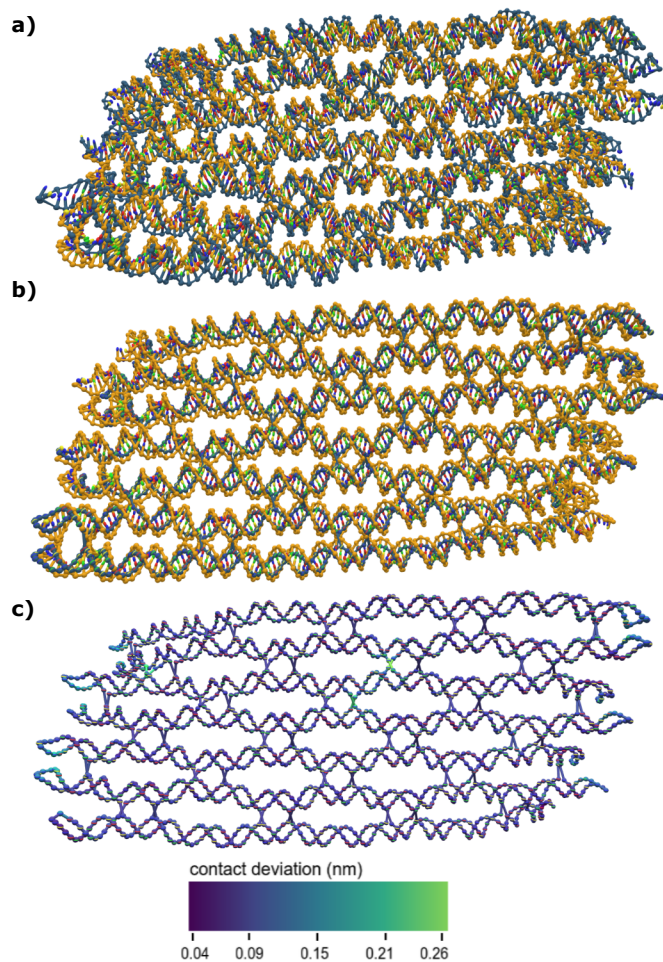


Figure 11. **Centroid structure and mean computed via multidimensional scaling.** **a)** The centroid structure (blue) observed during a simulation of a single-stranded RNA origami from⁸⁶ overlaid on the SVD-computed mean (yellow). This is the structure with the lowest RMSF to the mean structure. **b)** The mean structure as computed both by SVD (yellow) and MDS (blue). Because MDS does not preserve orientation data, the nucleotides are visualized simply as spheres at their center of mass, rather than having distinct base/backbone sites. **c)** The deviation in local contacts from the mean structure calculated in **b)**. This measure shows most of the structure to be homogeneously stable, with higher flexibility at helix ends and at junctions capable of sliding.

trajectory of the angle over the course of the simulation. Here, we show an example of the angle script again using the wireframe origami designs from¹⁰⁸. Each origami has a designed junction angle corresponding to the number of arms joined at each junction (Figure A2 in appendix A). Deviation from this designed angle is a measure of strain and how non-planar the structure is in simulation. This can be particularly revealing in combination with the mean structure, showing that an on-average flat structure has a significant degree of flexibility over the course of the simulation.

The tethered multi-fluorophore (TMF) structure from¹⁹⁹ was used as a demonstration of the distance script. This structure is used to measure binding kinetics through the large change in radius of gyration induced by binding and unbinding of compatible sequences near the ends of the double-stranded tether. End-to-end distance of the tether in both the bound and unbound states are shown in Supplementary Figure S3 in the Supplementary Materials. Knowing the end-to-end distance of this structure can be used in predicting the radius of gyration for various states of the structure, which is useful in corroborating experimental results.

2.3.2.3 Base Pair Occupancy

The hydrogen bonds defining Watson-Crick base-pairing are the single most important parameter defining DNA/RNA nanotechnology geometries. Since structures are designed towards a theoretical global free-energy minimum that maximizes hydrogen bonds, deviations from the designed structures point to regions of significant topological strain or that have found a kinetically trapped structure distinct from the intended design. OxDNA/RNA defines hydrogen bonds between base-paired nucleotides as a base-pairing potential between two base particle beads less than $-0.1k_bT$, about 10%

of the magnitude of the equilibrium value of the base pairing potential of a base pair in a duplex. The script compares the hydrogen bonds in a simulation with a provided list of pairs present in the intended design. The fraction of the configurations in which the intended bonds are formed are reported as an oxView overlay file, with color coding intensity corresponding to the fraction of the time where the bonds are formed. Bonding is considered 0 for nucleotides without designed complements.

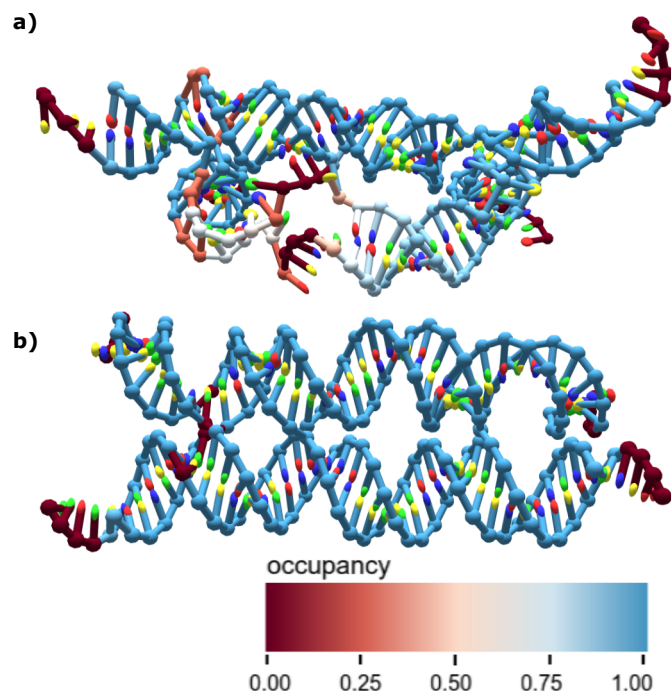


Figure 12. **Bond occupancy of an RNA tile.** **a)** The hydrogen bond occupancy during an oxRNA simulation, overlaid on a structure of an RNA tile. The structure was known to form poorly in the lab, and the simulation revealed significant strain on one duplex. The structure used here is the centroid of a trajectory based on the global fitting parameters discussed later. This was used as a visualization instead of the mean structure, as the unpaired duplex made the structure so flexible that the mean structure collapsed. **b)** The broken duplex from the structure in **a)** was extended by one base pair, and the simulation was re-run. Shown here are the hydrogen bond occupancies overlaid on the mean structure. In simulation, this significantly improved rigidity.

2.3.2.4 Principal Component Analysis of Nanostructure Motion Modes

Principal component analysis (PCA) is a common method for analyzing molecular simulation data that extracts the largest sources of deviation from the dataset²⁰⁹. First, using SVD, each configuration is aligned to a mean configuration (produced by either SVD or MDS) to remove rotations and translations from the data. Each nucleotide's deviation from its reference position in x- y- and z-coordinates is stored as its difference matrix. A covariance matrix is then constructed from the difference matrices, and the eigenvalues and eigenvectors are found through eigenvalue decomposition. These are then sorted in descending order with the highest eigenvalues representing the largest sources of variation in the structure. The eigenvectors generated by PCA represent an orthogonal basis for the reconstruction of every structure visited during the trajectory, and these reconstructions can then be used for clustering of distinct sampled conformations. Finally,



Figure 13. **Principal Component Analysis of a Holliday junction.** Principal component analysis of a Holliday junction visualized on oxView. Shown here is the top mode, which corresponds to a scissoring motion in the junction, with the arm ends having significantly higher average displacement than the crossover point.

the PCA script outputs a .json file for the oxView tool, which displays arrows on the structure corresponding to the sum of a user-defined number of components weighted by their respective eigenvalues.

To demonstrate the principal component analysis of DNA/RNA structures developed in this work, we ran it on a simulation of a Holliday junction (Figure 13). As one would expect for this structure, PCA reveals strong collective motion for the junction arms. The motion grows stronger at the ends of duplexes, while the crossover point shows little motion.

2.3.2.5 Unsupervised Clustering of Configurations Encountered in Simulation

The trajectories produced in an oxDNA/RNA simulation can be tens of gigabytes in size and explore an expansive amount of the configuration space available to the structure. In cases where multiple metastable states are visited during the trajectory, aggregate structural data, such as mean structures or base pair occupancy, might not be representative of the ensemble. This is due to the presence of these distinct metastable states. Here, we once again use the DBSCAN clustering algorithm²⁰⁰, as implemented in SciKit Learn¹⁹⁷, to automatically extract clusters of geometrically distinct structures from large trajectories and save each cluster as a separate file containing a collection of configurations that can be analyzed independently. The clustering algorithm can take any matrix of positions as an order parameter, whether that be principal component

coefficients of each configuration, or simply the distance between two particles. The DBSCAN algorithm is particularly good at clustering molecular simulation data where metastable states tend to form distinct clusters separated by a large energy barrier, such that observing transition states is relatively rare and multiple distinct densities are observed.

To demonstrate the utility of clustering using structural order parameters, we analyzed a simulation of an RNA tile structure (Figure 14), that is known to form two distinct structural isomers in experiment (unpublished results). In the simulation, two states were encountered, the correctly-folded structure, with three crossovers, and an unfolded structure, in which the paranemic cohesion²¹⁰ between two of the crossovers is lost, leaving essentially a Holliday junction (Figure 14, cluster 2). There are many potential order parameters that can be used to separate out these two structures. In this case, we chose to work with the most aggregate data: each configuration's position in principal component space.

The components produced by PCA (Figure A1 in Appendix A) represent a linearly independent basis for describing structures relative to the provided mean structure. This also means that every configuration used to compute the components can be mapped to a unique point in $3N - 6$ dimensional space. When applying DBSCAN to the positions of configurations in this space (described in detail in Appendix A), the distinct conformational isomers can be separated without further processing. In addition to the two expected configurations, this method also separated out another cluster (cluster 1 in Figure 14) of structures where the paranemic cohesion was correctly formed, but stacking was interrupted at the nick point, resulting in a non-planar kinked structure. The overlay in Figure 14a shows the fractional hydrogen bond occupancy compared with the original design. Of particular note is the large stretch of blue on the left side of cluster 2 where the bonds that form the paranemic cohesion are missing. The clusters were further analyzed using the angle script, identifying the distinct interduplex angles between each duplex in the structures (Figure 14b). These distributions show the fully formed structure (cluster 0) as having the lowest angle between the left duplexes in the first panel of Figure 9B and cluster 1 having a very defined angle between the central duplexes (Figure 14b, center).

2.3.2.6 Other Utilities

In addition to the specific structural measures discussed here, this package also contains additional utility functions for processing and displaying data. The first are two scripts that utilize the SVD superimposer from Biopython¹⁹⁶ for improving visualizations. The superimposing script takes multiple configuration files that share the same topology and returns them with their translations and rotations removed relative to the first

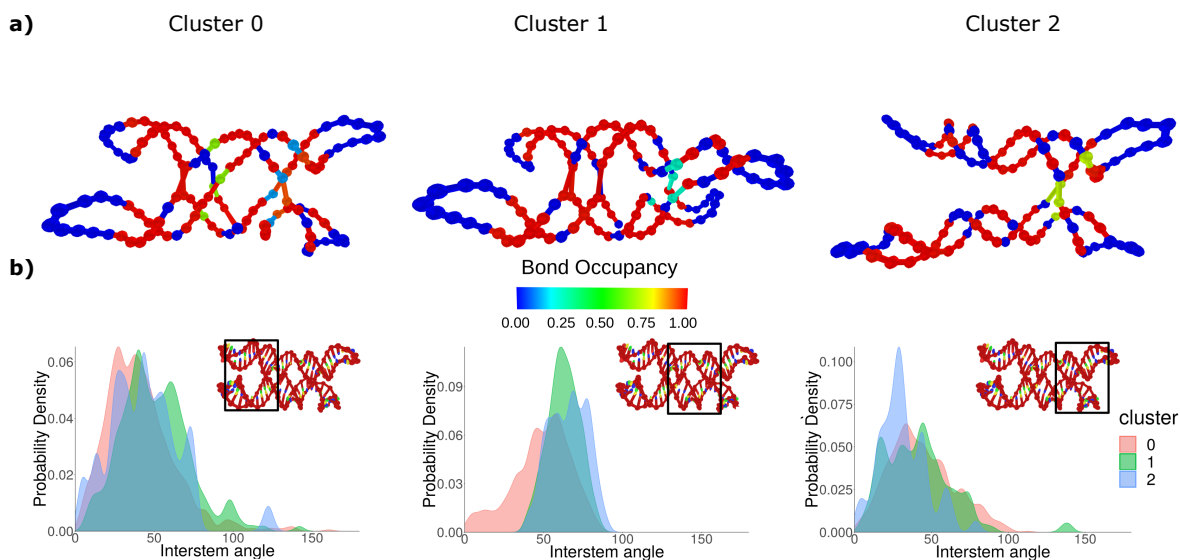


Figure 14. **Unsupervised clustering to isolate isomers of an RNA tile** **a)** The three clusters found in a simulation of a single-stranded RNA tile. The mean structure of each cluster was determined using MDS, and the hydrogen bond occupancy compared with the original design was used as an overlay. **b)** Histograms of the angles found in each cluster showing the distinct structures found in each cluster. The black frame on the tile snapshot indicates pairs of double-stranded RNA regions that were used to calculate the interstem angle.

configuration provided. We find this very helpful for comparing mean structures of similar designs or of the same design under different simulation conditions. There is also an alignment script, which takes a trajectory file and aligns all configurations to the first one in the file. This makes for a much smoother visualization experience when exploring trajectories in oxView or when making movies of a trajectory.

We have found the alignment scripts to be very useful for producing figures and movies (see Supplementary video 5 and Figure 11a) and for making comparisons between designs. These scripts are limited, however, by the need to align discrete units. Therefore, the structures must have the same number of particles in mostly the same position. Thus, the scripts are best used for comparing simulation conditions, changing sequences, and changing crossover positions in designs.

There is also a utility that reports the energy contribution of every interaction in the model. This has options of a text output to check specific values, as well as an oxView overlay showing the average energy of all nucleotides over the course of a simulation. Checking the base pairing or stacking interactions of specific nucleotides can be very helpful in identifying properties or defects in a given design. Additionally, we have found the visualization option useful for identifying excluded volume clashes during relaxations of large structures, as these cause extremely high total energies, which visually pop in oxView.

There are two further scripts that work with base pairs. One takes the current arrangement of base pairs

in the structure and generates either the designed pairs file used by the base pair analysis script, or an oxDNA mutual trap force file, which can be used to enforce a particular base pairing configuration during relaxation. This can be particularly helpful when relaxing multi-component structures edited in oxView, as the forces pulling stretched bonds back together can cause unwanted fraying of base pairs in otherwise stable structures. The second script converts oxDNA force files into a designed pair file. The Tiamat converter from¹⁷⁴ can produce force files as part of the conversion process, and this script can convert those force files into the format needed for the duplex angle script.

Finally, we provide a parallelization scheme for analyzing oxDNA trajectories. The parallelization module breaks down a trajectory into a number of chunks equal to the number of CPUs you have available, and uses the Pathos Multiprocessing library¹⁹⁸ to map trajectory chunks, CPUs, and functions. If the user has enough computational resources available, this facilitates analysis of even very large structures or long trajectories in a matter of minutes. The implementation of parallel functions is standardized across all scripts used here, and users are encouraged to follow the example given here in developing further analyses specific to their own designs.

Most of the analysis discussed fall into the class of tasks known as “embarrassingly parallel”, where there is no communication required between processes, and the final joining step is relatively easy. For all structure analysis algorithms described here, each configuration can be calculated independently of all the others. The only limitations to parallelization come from calculating split points in the trajectory and if a data trajectory is required, combining the outputs together in the proper order. As an example, we benchmarked parallelizing the computation of the mean structure of two structures: one with 423 nucleotides, and the other with 11 385. In both cases, runtime decreased by more than a factor of 10 when run on 30 CPUs compared with a single CPU, with diminishing returns past that point.

2.4 Discussion

We developed this collection of tools to remedy two gaps that we have perceived in the oxDNA software environment. First is the lack of an all-in-one visualizer that loads files within a reasonable timeframe, has a user-friendly UI, and performs edits on structures that could then be further simulated. All-atom simulations have such tools in the form of VMD, Chimera and PyMol. While tools exist to convert between all-atom and oxDNA formats, this is a cumbersome process that we felt could be remedied by the development of oxView. The use of hardware instancing allows oxView to load structures of unprecedented sizes and facilitates our

work on million-nucleotide oxDNA simulations of multi-origami structures. Furthermore, because oxView is built using the open-source 3D library Three.js, opens the possibilities for features from other Three.js projects to be added to oxView. For example, virtual reality oxDNA visualization was easily added by following the Three.js WebXR examples. Similarly, it is easy to export the visualized scene to other 3D formats, such as GLTF, for photorealistic rendering (Figure 6) or 3D printing (Figure A4 in Appendix A).

The features of oxView and simulation analysis tools are designed to help researchers in DNA and RNA nanotechnology to prototype *in silico* their structures, simplify the design and optimization process, and better understand the functioning of the designed structures. We demonstrated the utility and versatility of the visualization and analysis tools on multiple DNA and RNA nanostructure designs, ranging in size from hundreds to multiple thousands of nucleotides per structure. We also demonstrated that the tools can, in principle, handle structures of sizes over a million nucleotides.

These tools, particularly mean structure calculation and hydrogen bond occupancy, provide significant utility for iterative design of nanostructures. In many structures where unbounded growth is a goal, global curvature of the nanostructure due to subtleties in crossover placement is a significant bottleneck, that is difficult to solve using rational design principles. We have found that the curvature of mean structures calculated from oxDNA simulations (unpublished results) is a good predictor of lattice formation in the laboratory. We also note that mean structures are the best proxy for comparing simulations with cryo-EM structures, which have become important characterizations for 3D nanostructures in the nucleic acid nanotechnology field.

Hydrogen bond occupancy is a good proxy measure for the amount of stress built up in a structure. Even with the speed and level of coarse-graining that oxDNA provides, modelling assembly pathways for large structures remains out of reach for all but the most ambitious simulations¹⁶⁶. Because of this limitation, we perform simulations with the assumption that the structure forms as designed, and initiate the simulation with all hydrogen bonds present. Designed pairs that become unbonded or find different partners, particularly at junction points, are a good indication for points in the design that are stressed and would benefit from iterative design. In general, we found that successfully published structures had near 100% bond occupancy, while those that were proving difficult to obtain in the lab had regions with low occupancy.

We demonstrated the functionality and versatility of these tools by applying them to a range of DNA and RNA nanostructures, such as DNA and RNA origamis, as well as optimizing and analyzing an RNA tile.

All software discussed here is open-source and freely available through our GitHub under the GNU Public License. Pull requests, bug reports and feature suggestions are welcome, as we hope that these will pro-

vide fundamental support long into the future. All tools that were introduced here are documented on their respective GitHub repositories, with examples of use reproducing the figures in this paper.

2.5 Data Availability

The oxDNA simulation code is available online on the oxDNA webpage dna.physics.ox.ac.uk. OxView is available as a web-based application on github.com/sulcgroup/oxdna-viewer. The analysis package can be downloaded from github.com/sulcgroup/oxdna_analysis_tools.

2.6 Acknowledgements

We would like to thank R. Hariadi, C. Simmons, and X. Qi for the design files from their previous publications used as examples in this work. We would like to thank members of the Šulc and Yan labs, specifically H. Liu, J. Procyk, and L. Chen for their feedback and beta-testing of the tools presented here.

2.7 Funding

This work was supported by the National Science Foundation under grant No. 1931487. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska–Curie grant agreement No 765703. The EU funding concerns the contribution of JB.

2.7.1 Conflict of Interest Statement.

None declared.

Chapter 3

ONLINE RESOURCES FOR CHARACTERIZING DNA AND RNA NANOSTRUCTURE DESIGNS

This chapter was published in **E. Poppleton**, R. Romero, A. Mallya, L. Rovigatti, P. Šulc, **OxDNA.org: A public webserver for coarse-grained simulations of DNA and RNA nanostructures** *Nucleic Acids Research* 49(W1), e72 (2021)

Abstract

OxDNA and oxRNA are popular coarse-grained models used by the DNA/RNA nanotechnology community to prototype, analyze and rationalize designed DNA and RNA nanostructures. Here we present oxDNA.org, a graphical web interface for running, visualizing and analyzing oxDNA and oxRNA molecular dynamics simulations on a GPU-enabled high performance computing server. OxDNA.org automatically generates simulation files, including a multi-step relaxation protocol for structures exported in non-physical states from DNA/RNA design tools. Once the simulation is complete, oxDNA.org provides an interactive visualization and analysis interface using the browser-based visualizer oxView to facilitate the understanding of simulation results for a user's specific structure. This online tool significantly lowers the entry barrier of integrating simulations in the nanostructure design pipeline for users who are not experts in the technical aspects of molecular simulation. The webserver is freely available at oxdna.org.

3.1 Introduction

The field of nucleic acids nanotechnology uses DNA and RNA molecules as basic building blocks to construct nanoscale structures and devices. DNA and RNA have been chosen due to their programmability, which exploits the complementarity between corresponding bases (A-U/T, C-G) to design target nanostructures as overall free-energy minima of systems composed of self-assembling single DNA or RNA strands. Over the past four decades since its inception³⁵, the field has led to the production of increasingly larger and more complex self-assembled structures with applications that include synthetic biology²¹¹, nanopatterning¹⁷⁸, nanophotonics to drug delivery⁵⁶, diagnostics²¹², immunotherapy²¹³, and vaccine development²¹⁴. Experimental techniques, such as fluorescent labeling, AFM and cryoEM are typically used to characterize the structures. However, the resolution of the experiments is limited, and most structures are typically designed empirically through a trial-and-error procedure, until the desired shape or structure property is achieved, which is time-consuming and costly.

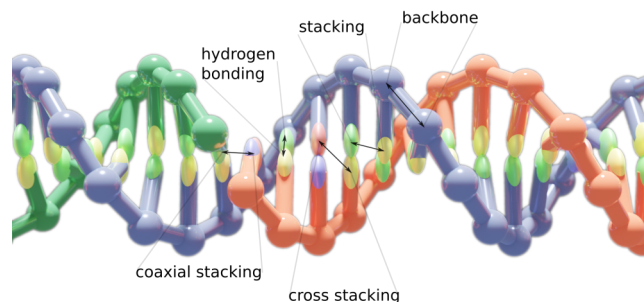


Figure 15. **The oxDNA model** is a coarse-grained one-bead-per-nucleotide model of DNA with empirically derived potentials between the beads which replicate the physical, mechanical and thermodynamic properties of double- and single-stranded DNA molecules. The electrostatic repulsion between nucleotides is implemented using Debye-Hückel potential. The oxRNA model uses similar nucleotide-level representation for RNA molecules as oxDNA does for DNA.

An alternative is provided by computer simulations, which can offer detailed insight into the function and properties of DNA and RNA nanostructures. Using atomistic resolution models faces challenges due to the size of the systems (ranging from hundreds to tens of thousands of nucleotides), and long timescales involved in assembly as well as equilibrium states sampling for typical nanostructure designs. Hence, coarse-grained models, which use simplified representations that group multiple atoms into a single particle with effective parametrized interactions, have become increasingly popular to study nanostructures^{144,215}. Finite element-based computational studies of mechanical and structural properties of nanostructures^{126,127} have also been developed. Among the most popular tools in the past few years have been the oxDNA and oxRNA models for DNA and RNA nanotechnology modeling^{140,141,143,159}. They represent each nucleotide as a single rigid body with empirically designed interactions that are parameterized to reproduce basic structural, mechanical and thermodynamic properties of both single-stranded and double-stranded DNA (Fig. 15). Where available, the models have been found to be in good agreement with experimental data and have been used in over 130 articles in the past ten years. Applications range from studies of biophysical properties of DNA and RNA to studies and rationalization of the function of DNA nanodevices, probing nanostructure design and simulations of their assembly^{54,145,146,166,176,184,189,191,192,215-217}.

However, there is a steep learning curve in using the simulation code. It requires access to GPU-equipped servers, knowledge of command line environment and practical experience in setting up and evaluating molecular dynamics (MD) simulations, which typically requires at least basic programming expertise as well. One of the most common use cases of oxDNA and oxRNA is to prototype and test novel nanostructure designs in

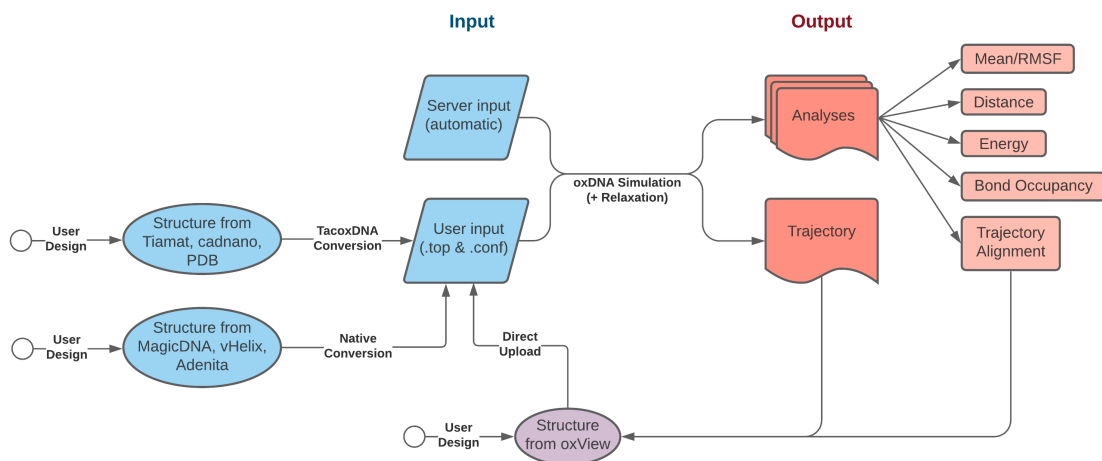


Figure 16. **Workflow of the oxDNA.org pipeline**

equilibrium sampling simulation. To ease the use of the tool and also make it accessible to a broader user base, we introduce here a public webserver, oxDNA.org, with a GUI that integrates automated simulation setup and subsequent evaluation in an intuitive user-friendly environment. We provide here the description of the simulation setup and analysis workflow, as well as information about the models and the file formats used. Further tutorials and examples of how to use the server are provided online.

3.2 Methods

3.2.1 Server Data Processing

OxDNA.org runs the public release of the oxDNA simulation code that implements the oxDNA2 and oxRNA models. It facilitates user interaction via a graphical user interface (GUI) that automatically generates parameters for the simulation and simplifies post-processing. The data workflow used by oxDNA.org is shown in Fig. 16. The details about the parametrization of the models and the computational implementation can be found in previous work^{140–143,159}. The oxDNA.org webserver brings together, for the first time, structure relaxation, simulation, visualization and post-processing in a single GUI environment. Users provide the DNA and RNA structure input files in the oxDNA format and are presented with the option to choose a limited number of simulation parameters:

- Input files – An oxDNA configuration and topology file pair that define the structure that will be simulated

- Job Name – The name you would like to give the job
- Interaction Type – Whether to use the oxDNA2 or oxRNA force field when running the simulation
- Salt Concentration – The monovalent ion concentration in molar. OxDNA uses the Debye–Hückel electrostatic model to implicitly model reduction in backbone electrostatic repulsion due to the presence of monovalent ions. Magnesium interacts with DNA and RNA in a non–uniform, site–specific manner and therefore is not included in the oxDNA model. Previous studies¹⁶⁵ have shown that high monovalent ion concentration in the simulation produce results very similar to those found in experiments containing standard levels of magnesium ions.
- Steps – The number of steps to run the simulation (running time = steps · timestep)
- Temperature – The temperature of the simulation (set with an Andersen–like thermostat).
- Relaxation – If checked a relaxation protocol will be run prior to the production simulation. The protocol comprises a Monte–Carlo (MC) relaxation and an MD relaxation.

The following additional parameters are available for the relaxation protocol:

- MC Steps – The number of steps to run the MC relaxation for. This is mostly used to relax overlapping particles and stretched bonds. If there are not many of these in the input structure, this can be shorter than the default.
- MD Steps – The number of steps to run the MD relaxation for. If structures do not completely relax, this step should be made longer.
- MD Timestep – The timestep of the MD relaxation. If structures fail to relax or explode during the relaxation, it can be helpful to lower this value.

The following parameters are also available in the “advanced parameters” section for the production simulation:

- Backend – Run using the CPU or CUDA (GPU) backend. For structures over a couple hundred nucleotides the CUDA backend will be significantly faster. The server is equipped with 8 GPUs and 20 CPU cores, so small structures should be run on CPU to leave GPU capacity for larger structures.
- Simulation Timestep – The timestep of integration for the MD simulation. This number should not be set higher than 0.003 to avoid numerical instability.
- External Force File – The user may upload an external force file to add additional forces, for example, to pull two separated strands into proximity to facilitate binding. The external force file can be produced using the oxView visualization and analysis tool¹¹⁴ by selecting which nucleotides should be paired

together. Alternatively, the force files can also be used to pin a specific nucleotide to a given position, or to introduce a 2D plane in the simulation box, as documented in dna.physisc.ox.ac.uk.

- Average Sequence Model – By default, oxDNA and oxRNA uses the same (averaged) sequence strength for A–T(U) and C–G base pairs and for stacking interactions. Switching this option will run the model with sequence–dependent strengths for A–T and G–C bonds (or A–U, G–C and G–U for RNA) as well as with sequence–dependent stacking interaction strengths.
- Mismatch Repulsion – The oxRNA force field is known to over–stabilize mismatches between paired segments. This introduces an additional repulsion force between non–complementary bases to reduce such incidents.
- Print Conf Interval – The frequency with which the simulation will print its current configuration to the trajectory file. The default number was chosen to obtain configurations that are, on average, uncorrelated. OxDNA limits data output to 5 MB per second, so the maximum print rate depends on the size of the structure.
- Print Energy Interval – The frequency with which the simulation will print its current energy to the energy file.

The default parameters were chosen to provide a good balance between runtime and sampling and in most cases will result in a trajectory with uncorrelated energies between subsequent configurations. The default parameters will run a DNA simulation at 20° C and 1 M sodium concentration for $1e9$ MD steps with a timestep of integration of 0.001 simulation time units. Note that, due to the inherent nonlinearity in coarse–graining, an exact time correspondence with experiments cannot be established¹⁶⁶. Based on direct unit conversion the default running time corresponds to $3.03 \mu s$. However, the choice of the diffusion coefficient in the simulation makes diffusion 100 times faster than what has been measured experimentally. Hence, the $1e9$ steps corresponds roughly of up to 1 ms in real time. However, different processes might scale with different ratios.

For structures exported from different DNA/RNA nanostructure design software, a relaxation step is often required to achieve a physical configuration prior to starting the oxDNA simulation. For example, the respective parts of the structures might be drawn on a lattice and the nucleotide positions in the design interface might violate the possible length constraints imposed by covalent bonds. Other common problem with structures directly exported from different design tools might be that some nucleotides are positioned too close to each other, leading to steric clashes. An MD simulation cannot be directly started from such configurations, since the large forces due to the unphysical conformations would lead to numerical instabilities. OxDNA.org hence

implements the relaxation scheme described in¹⁵⁸ where a short Monte–Carlo simulation is first performed to reduce excluded–volume clashes and shorten stretched bonds. The structure is then relaxed using a longer MD simulation with a highly–coupled Bussi–Donadio–Parrinello thermostat¹⁵⁷ and a modified backbone potential which reduces the possibility of numerical instabilities. The default relaxation parameters on the webserver are very generous in order to facilitate most user submissions. Structures that are already in a near–physical state can be run faster by reducing the number of steps in the relaxation, while some structures that are very far from their expected configuration may require more aggressive methods beyond the scope of oxDNA.org such as rigid–body dynamics¹¹⁴ (which can be performed in the oxView windows on oxDNA.org, however it is more convenient to use the full–screen version on the main website), mrDNA¹⁴⁴, or interactive relaxation using oxServe²¹⁸.

Simulations of a full–sized DNA origami (approx. 10 000 nucleotides) take about 3 days to run on oxDNA.org with the default parameters, with runtime scaling approximately linearly with the total number of nucleotides. Users are allowed to submit up to 4 simulations at a time, and the trajectories are kept on the server for one week after completion. The trajectories are stored in a compressed 7zip format, however for a origami–sized structure the files will still be approx 3 GB, so ensure sufficient time to download the results. The server is currently equipped with 8 NVIDIA RTX 2080 Ti GPU cards with plans to expand capacity in the future.

3.2.2 Software

The web frontend/backend uses a Flask (Python)/Angular 1.8 (JavaScript)/Bootstrap(CSS+HTML) stack. The main code of oxDNA is written in C++ and CUDA and can be downloaded from (dna.physics.ox.ac.uk). OxView (sulcgroup.github.io/oxDNA-viewer/) is a single–page Three.js (JavaScript) application used for visualization and editing of oxDNA structures. Analysis is performed using the Python–based `oxDNA_analysis_tools` package (github.com/sulcgroup/oxDNA_analysis_tools). Examples of outputs are demonstrated in the following section.

A

You are now creating a job for processing on our servers.

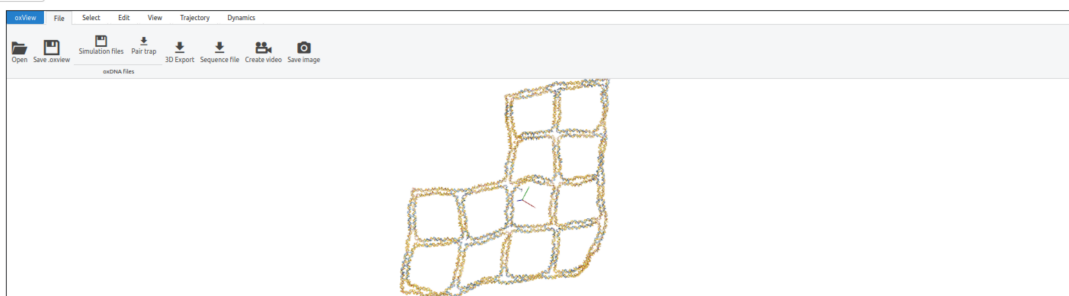
Server Status

Total jobs currently running: 8
Total jobs in queue: 14

Input Files

Please select 2 input files (configuration and topology) for the job

Browse... 2 files selected.



Parameters

Job Title:

Interaction Type: DNA RNA

Salt Concentration:

Steps:

Temperature Units: Kelvin Celsius

Temperature:

Relaxation: Needs Relax

[View oxDNA parameters](#)

B

View Jobs

Server status:
Total jobs currently running: 9
Total jobs in queue: 8

Your simulations:
Here is a list of the jobs you have currently running, pending or finished under your account. Please be aware trajectory files get automatically deleted one week after the job finishes. To access analysis tools, click on your job name.

Job Name	Status	Started	Input	Log	Topology	Initial Configuration	Last Configuration	Trajectory	Cancel	Delete
Example	Running	2/16/2021, 3:49:47 PM	view download	view download	download	view download	view download	download	<input type="button" value="Cancel Job"/>	<input type="button" value="Delete Job"/>

Figure 17. **Submitting a job on oxDNA.org.** The Job submission interface showing the uploaded structure in oxView and the parameter setting (A). The job status page where users can check the status of their jobs as well as download the original input files as well as the output files and logs. Clicking on the job name will take the user to the analysis page for that job (B).

3.3 Results and Discussion

3.3.1 Input Files for the Server

The job submission form on oxDNA.org (Fig. 17) requires two files: configuration and topology. Both files are described in detail in the oxDNA documentation on dna.physics.ox.ac.uk. Briefly, the configuration file header has three lines that contain, respectively, the timestep, simulation box size, and total, potential and kinetic energy per particle for the given configuration. Then, the file contains one line per nucleotide. Each line contains nucleotide position, the normal particle orientation vectors \mathbf{a}_1 and \mathbf{a}_3 of the reference frame of the nucleotide, and the velocity and angular velocity vectors. The topology file contains connectivity information defining which particles are connected together to form strands as well as the nucleoside identity of each particle. The first line of the topology file shows the total number of nucleotides and strands. Then, for each nucleotide, its corresponding line in the topology file lists the strand id that the nucleotide is part of, the base identity (A,C,G, T or U for RNA), the id of the nucleotide's 3' neighbor along the strand and then the id of the nucleotide's 5' neighbor. The nucleotides in the topology file are listed 3' to 5' (note the backwards convention) for each respective strands. If a nucleotide does not have a 3' or 5' neighbor, the id of the neighbor is listed as -1. The ids of the strands start from 1, the ids of the nucleotides start from 0.

There are a variety of popular design tools in the DNA/RNA nanotechnology field^{94,97,106,111,112} that can be used to create a starting configuration for an oxDNA simulation. Some (like Adenita, MagicDNA or vHelix) have built-in exporters to the oxDNA format while others (caDNAno, Tiamat, or PDB format) can be converted using TacoxDNA (tacoxDNA.sissa.it)¹⁷⁴. OxDNA files can be edited using the oxView tool, either on the main website (sulcgroup.github.io/oxDNA-viewer/) or in the small viewer window when submitting jobs on oxDNA.org. Users may optionally include an external force file which defines an external potential that acts on certain particles in the simulation. Most commonly used are mutual traps, which add an external spring potential of a given stiffness and equilibrium length between two particles. These files can be generated using oxView or oxDNA_analysis_tools or can be manually written based on the template provided in the documentation.

When submitting a job, users can either sign up for an account to keep track of all their submissions in one place or submit anonymously, in which case the user should bookmark the job output link for later access. Users who create an account may opt-in to emails when jobs complete and will receive email reminders about stale files that will soon be removed from the server.

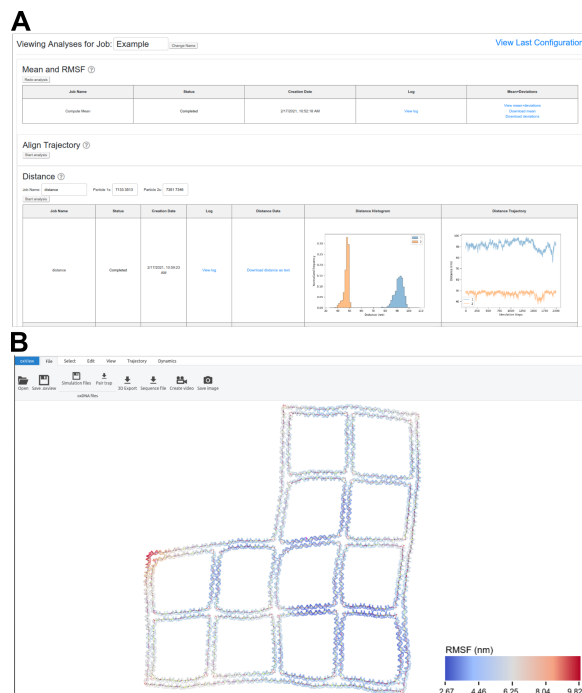


Figure 18. **Analyzing a completed job.** The job analysis interface showing a few of the options available to users when their simulation finishes. **(A)** In this example, a wireframe DNA origami from¹⁰⁸ was simulated and, when finished, the mean structure with RMSF and the end-to-end distances of two of the edges were calculated. Clicking on the “view” options will open the structure in a separate oxView tab allowing the user to interactively explore their results. **(B)**.

3.3.2 Server Output Result

There are two pages of outputs from simulation jobs run on the server (Fig. 18). The first is a summary table of all jobs run by the user which contains links to download the initial configuration and topology files submitted, as well as the simulation input file and job logs generated by the server. The user can further view or download the last configuration output by the simulation and download a zip archive containing the entire simulation trajectory. By clicking on the job name, the user will be taken to the analysis page where they will find many options for post-processing of their simulation trajectory. This page is a GUI implementation of many of the scripts found in `oxDNA_analysis_tools`. The following scripts are available through the GUI:

- Mean and RMSF – Uses single value decomposition (SVD) alignment to calculate the mean position of every nucleotide, then again analyzes the trajectory to get the root mean squared fluctuation (RMSF) of each particle from its mean position.
- Align Trajectory – Uses SVD to align all frames in the trajectory to the first. Produces a clearer view

of fluctuations when viewed or converted to a movie using oxView. This creates a trajectory file about 2/3 the size of the original, so is recommended prior to downloading.


- Distance – Calculates the distances between two lists of particles (the first particle id in “particles 1” is compared with the first particle id in “particles 2” etc.) Results are provided as both a histogram and line plot format in addition to as a text file.
- Energy – Creates plots out of the simulation energy file showing the average potential energy per particle. Fluctuation around a constant value is the simplest way to determine whether or not a structure has been properly equilibrated.
- Bond Occupancy – Calculates the percentage of trajectory frames in which the bonds present in the first configuration are subsequently present. This is useful for prototyping structures, as improperly designed regions will be unstable and prone to breaking base pairs.
- Duplex Angles – This is split into two sections, the first calculates the orientation of all duplexes in the structure. The second takes as input the starting nucleotide ids of duplexes and uses the output file from the previous script to calculate the angle between duplexes that start with those particles. The input and output format are the same as the distance script.

“Mean and RMSF”, “Energy Plotter”, and “Bond Occupancy” give a good summary of how the structure behaves and what fluctuation modes the structure encountered during the simulation. Base pair occupancy in particular is something that rapid equilibrium structure prediction algorithms such as CanDo¹²⁶ and SNUPI¹²⁷ cannot predict. Distance and Angles serve to ask specific questions about how regions of the structure behave relative to one another, allowing the user to accurately discern parameters such as ideal locations for FRET pairs or explore 3D curvature that is not visible on AFM images. All results from the server can be downloaded by clicking the download link.

The webserver has been tested on all major browsers and operating systems (Table 1). The original oxDNA code is supported only for Unix-based systems, so this improves accessibility to those using Windows environments.

3.4 Conclusion

OxDNA.org simplifies the oxDNA/oxRNA simulation pipeline and makes a tool previously limited to Unix command line execution usable to non-expert users via a GUI. OxDNA has long been a popular tool for the



Linux	Yes	Yes	N/A	N/A
Windows	Yes	Yes	N/A	Yes
OSX	Yes	Yes	Yes	N/A

Table 1. **Browser Compatibility** oxDNA.org works in all major browsers and operating systems.

prototyping of DNA and RNA nanotechnology structures, and we hope that the simplification of the interface and the implementation of automatic relaxation will allow this tool to find wider adoption among traditionally experiment-only groups and that simulation-probing will become a standard step within the nucleic acid nanostructure design and characterization process. The webserver implementation and scripts are also freely available under GNU Public License at github.com/sulcgroup/oxdna-web.

3.5 Acknowledgements

We thank the members of the Yan and Šulc groups at ASU who participated in the server beta-testing, in particular to Michael Matthies, Jonah Procyk, Hao Liu, Joel Joseph, Shuchi Sharma, Kai Yin, Jeffrey Luo, Jacob Baca, Christopher Persson, and Swarup Dey. We further thank Doye and Louis group members from Oxford University for their help with debugging the deployed server.

3.6 Funding

This work was supported by the National Science Foundation under grant no. 1931487.

Chapter 4

USING OXDNA SIMULATIONS TO CHARACTERIZE THE EFFECTS OF DESIGN CHOICES ON A LEAF-SPRING NANOENGINE

This chapter previously appeared in preprint form M. Centola, **E. Poppleton**, M. Centola, J. Valero, P. Šulc, M. Famulok, **A rhythmically pulsing leaf-spring nanoengine that drives a passive follower** *BioRxiv* (2021)

Abstract

Molecular engineering seeks to create functional entities for the modular use in the bottom-up design of nanoassemblies that can perform complex tasks. Such systems require fuel-consuming nanomotors that can actively drive downstream passive followers. Most molecular motors are driven by Brownian motion, but the generated forces are scattered and insufficient for efficient transfer to passive second-tier components, which is why nanoscale driver-follower systems have not been realized. Here, we describe bottom-up construction of a DNA-nanomachine that engages in an active, autonomous and rhythmical pulsing motion of two rigid DNA-origami arms, driven by chemical energy. We show the straightforward coupling of the active nanomachine to a passive follower unit, to which it then transmits its own motion, thus constituting a genuine driver-follower pair. Our work introduces a versatile fuel-consuming nanomachine that can be coupled with passive modules in nanoassemblies, the function of which depends on downstream sequences of motion.

4.1 Introduction

Active mechanical motion of nanoscale objects is paramount for the bottom-up construction of bio- or technomimetic nanomechanical machines²¹⁹⁻²²⁴ that can perform tasks like pumping²²⁵, walking²²⁶, transduction or sensing of molecules or signals⁹, or any process involving motion^{227,228}. Both in the nano- and in the macroscopic world these processes require fuel-powered engines that perform periodically repeating rhythmic motion. Impressive examples of synthetic pumping, rotating, or moving fuel-driven nano-devices exist²²⁹⁻²³³ but the creation of engines that generate active rhythmic motion at the nanoscale, driven by chemical fuel, remains challenging²³⁴⁻²³⁶. Here we report a bio-hybrid nanoengine (NE) that rhythmically pulses by consuming nucleoside triphosphates as fuel to build up potential energy stored as spring-tension in a compliant flexure mechanism, followed by active relaxation. The device consists of two rigid DNA origami-arms connected in

a V-shape by a bendable flexure at the pointed end, and a double-stranded DNA (dsDNA) sequence that spans the diverging ends²³⁷. One origami-arm has a T7-RNA polymerase (T7RNAP) covalently attached near a T7 promoter sequence in the dsDNA, allowing its use as a transcription template to create traction²³⁸⁻²⁴⁰. Upon transcription, the stationary T7RNAP pulls the dsDNA like a rope and drags the opposing origami-arms towards itself, which builds up spring tension at the compliant flexure that is released when the polymerase reaches a termination sequence near the opposite origami arm allowing the structure to return to its equilibrium position. This design leads to a continuous rhythmic flapping motion the frequency of which can be tuned by slight changes in the transcribed dsDNA sequence. Coarse grained molecular dynamics simulations illustrate the mechanical properties of various nanoengine designs and confirm our experimental results. In a prototypical application we show that the engine acts as a mechanical driver that can be coupled to a passive follower unit to which it then can actively transfer its motion, opening ground-breaking opportunities for its future use to drive more complex nanomachines, similar to the balance wheel in a watch or in Leonardo da Vinci's self-propelled cart.

4.2 Design of the Nanoengine

The two 60 nm long origami-arms are formed by 18 helix bundles (18hb) that are arranged in a honeycomb structure to form rigid and straight arms (Figure 19a, for exact sequences and CadNano map see Supplementary Dataset S1 and Supplementary Dataset S2). They are connected by six 28 nm long dsDNA helices arranged as a 12 nm wide sheet that can be bent to serve as the compliant leaf-spring, as shown in previous work^{237,241}, and by six single-stranded (ss) DNA strands that ensure the formation of the bent V-shape by being overall shorter than the double-stranded compliant area (Figure 19a-e; Figure B1a in appendix B). A 154 nucleotide (nt) long transcribable dsDNA template strand (Figure 19a,c, grey; Figure B1b in appendix B) spans the origami-arms and is firmly connected to each of them at 30 nm distance (Figure 19a; Figure 19e, red dots) from the leaf-spring ends. One of the origami-arms contains a sequence with a 5'-chloroalkyl group attached near the template dsDNA (Figure 19e, yellow dot; Figure B1c in appendix B), to which a HaloTag-T7RNAP fusion protein (Figure 19a-c, orange-blue; Figure B1b,d,e in appendix B) couples covalently at its HaloTag²⁴² (HT) subunit (Figure B1f in appendix B). The dsDNA template contains a T7 RNA polymerase promoter region (yellow, Figure B1a,g in appendix B) and a sequence that, once transcribed, binds to a molecular beacon (green, Figure B1a,g in appendix B) to monitor the amount of RNA generated during transcription. The proximity of the HT-attachment site and the dsDNA template (Figure 19e, yellow and red

dots) helps the T7RNAP bind to the T7 promoter sequence in the template (Figure B1a,b in appendix B, yellow) and facilitates initiation of transcription in the presence of NTPs until it reaches the terminator sequence at the opposite end of the dsDNA template (Figure B1a,b in appendix B, red). Moreover, the opposing origami-arm contains four biotinylated sequences that protrude the arm at the outer part of the structure to which four streptavidin proteins can be attached to unambiguously identify each arm by atomic force microscopy (AFM, Figure 19d,f) or transmission electron microscopy (TEM, Figure 19g). The protruding biotin residues also can potentially serve as attachment points to anchor the NE on surfaces or to other origamis. These features were designed so as to operate the NE autonomously and continuously in a rhythmic opening/closing cycle (Figure 19h) starting from the open equilibrium conformation of the structure, in which the T7-promotor is bound by the HT-T7RNAP to begin transcription (Figure 1h, 1). The pulling of the template DNA through the immobilized T7RNAP closes the origami structure while building up spring tension in the compliant segment of the structure that generates a counteracting force (Figure 19h, 2). Once the terminator sequence has been reached, the polymerase releases the template dsDNA and the structure opens to its equilibrium conformation (Figure 19h, 3) and is ready to begin a new cycle.

The purity of the origami was analysed by gel electrophoresis and the structural integrity of NEs was confirmed by AFM (Figure 19f), and TEM (Figure 19g). In the AFM images of NEs that are not engaged in transcription the dsDNA template is clearly visible in between the origami-arms. The TEM images also show the HT-T7RNAP (Figure 19g, right panel, blue arrow) and the three streptavidin-tags (green arrows). AFM distance measurements by height profiling confirm the designed distance of 21 nm between the streptavidin tags and the height of the origami-arms, respectively (Figure B2c-e in appendix B). We also analysed NEs during transcription by TEM (Figure 19i,j). The TEM image (Figure 19i, right) shows that the DNA template strand exists in a strained state during the initial phase of transcription, as shown in the 3D model (left). Examples of NEs in the final phase of transcription show the origami-arms in a "closed" state (Figure 19j, right), as shown in the 3D model (left).

4.3 Molecular Dynamics Simulations

Coarse-grained simulations using the oxDNA model¹⁴⁰⁻¹⁴³ of different NE designs were performed to further characterize the impact of our design choices on mechanical properties (Figure 20a). In addition to verifying agreement with experimental results, there were three questions that we explored using the higher

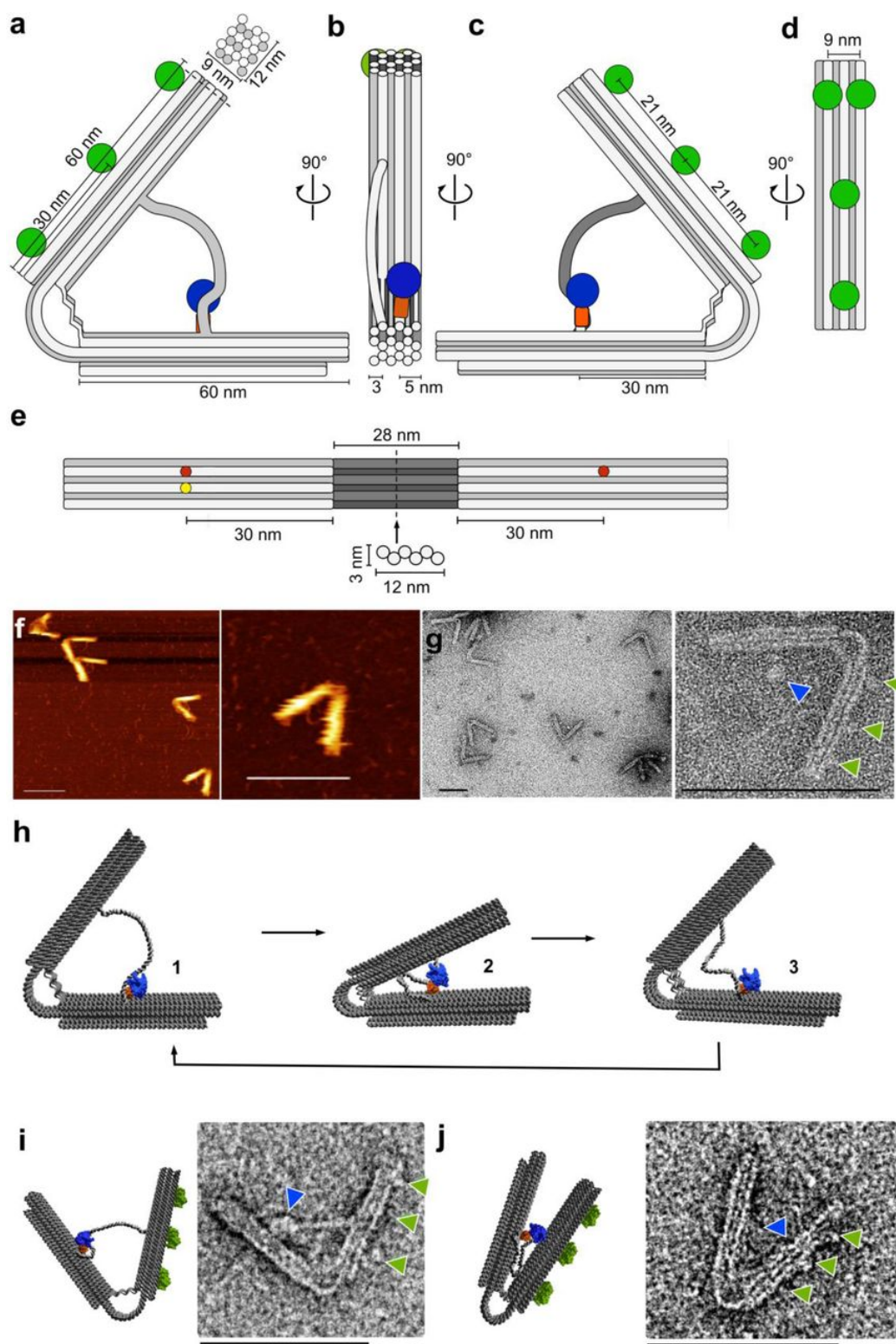


Figure 19. (Previous page.) **Design and dimensions of the DNA leaf-spring nanoengine.** (a) Cross-section of the leaf-spring NE. Dimensions of the stiff origami-arms. Green circles: attachment sites for streptavidin binding; blue circle: T7RNAP-part of the HT-T7RNAP fusion protein. Top panel: arrangement and dimensions of the 18-helix bundle that forms the origami-arms, (b) cross section of the 90° left turn of the cross section shown in (a); orange: HaloTag (HT), blue: T7RNAP, (c) cross section and dimensions of the 90° left turn of the cross section shown in (b), (d) cross section and dimensions of the 90° left turn of the cross section shown in (c), (e) cross section and dimensions of the origami-arms that flank the 28 nm long leaf-spring helices (dark grey) that are arranged in a honeycomb structure (bottom panel), red dots: attachment sites of the dsDNA template strand, yellow dot: attachment site of the HT-T7RNAP. (f) Characterization by Atomic force microscopy (AFM) of the leaf-spring NE. Overview (left) and detailed image (right) of the NEs in alternating contact mode in air on a poly-L-ornithine-functionalized mica surface, (g) Transmission electron microscopy (TEM) of the NE in negative staining. Overview (left) and detailed image (right) of the NEs. Green arrows: streptavidin molecules bound to biotin-modified staples protruding from one of the origami-arms opposite to the location of the HT-T7RNAP fusion protein (blue arrow), (h) Full opening and closing cycle of the compliant mechanical structure. **1** in the open structure the dsDNA template is bound by the immobilized HT-T7RNAP fusion protein and transcription begins. **2** upon transcription, HT-T7RNAP pulls the opposing origami-arm towards itself forcing the structure to close. **3** When the terminator sequence is reached, the T7RNAP releases the dsDNA template linker, which causes the structure to snap open to its equilibrium conformation. The T7RNAP can initiate the next closing cycle. (i) Example of the NE engaged in transcription. Blue arrow: HT-T7RNAP, green arrows: streptavidin, (j) Example of the NE engaged in transcription. Blue arrow: HT-T7RNAP, green arrows: streptavidin. All scale bars: 100 nm.

degree of resolution and control afforded by molecular dynamics: (1) the relative rate of opening and closing the structure, (2) The contribution of the ds bridge to the re-opening rate, and (3) the effect of scaffold sequence across the single-stranded regions of the flexure. Notably, the simulations showed that the largest influence on mechanical properties came from transient, sequence-dependent secondary structures formed by the single-stranded domains in the flexure. Disallowing formation of these secondary structures ("no structure": all structures labelled as NS in Figure 20b) increased both the equilibrium angle and the stiffness of the (n)NE (Figure 20b). Conversely, allowing the formation of secondary structures reduces the equilibrium angle to roughly the same value, irrespective of variations in the transcribable sequence (NE_NB, NE, and nNE in Figure 20b). This effect is particularly pronounced in the NE-structure lacking the transcribable DNA sequence ("no-bridge": labelled as NB in Figure 20b), where the difference in equilibrium angle between NE_NB and NE_NS_NB is 25°. This result suggests that the secondary structures in the hinge region are the largest determinant of equilibrium angle and stiffness (Figure B3a in appendix B); however, the presence of the transcribable sequence imposes a hard upper limit on the opening angle of the structure.

In the absence of the transcribable dsDNA and secondary structures in the hinge region, the DNA leaf spring can stretch out and reach an equilibrium angle of 80° with a standard deviation of 9°. When the transcribable sequence is introduced, NE_NS and nNE_NS have comparable equilibrium angles of 71 ±6° and 74 ±5°, respectively, both with a negatively skewed distribution caused by the upper limit imposed by the transcribable sequence. When secondary structures are permitted, the scenario changes drastically: now

NE_NB, nNE, and NE have highly comparable equilibrium angles of $55^\circ \pm 8^\circ$, $58^\circ \pm 7^\circ$, and $54^\circ \pm 7^\circ$ respectively (Figure 20b), implying that the effect of the transcribable dsDNA on the equilibrium angle is negligible. These results suggest that the hairpins formed by the single stranded sequences decrease the length and flexibility of the hinge region, influencing both the overall shape and rigidity of the origami. The structures in the hinge region are only slightly affected by temperature changes. Increasing the simulation temperature from 23 °C to 37 °C only marginally increases the angles observed in the simulated structures. (Figure B3b in appendix B).

To test whether the predicted effects of the simulations can be confirmed by experimental results, we analysed the angle distribution using TEM images of a construct lacking the transcribable dsDNA (NE-NB) and compared it with that of NE (Figure B3c,d in appendix B). We observed that the two populations have the exact same distribution ($p = 0.6$), suggesting that the secondary structures predicted by the simulations are indeed present in experimental conditions and have an impact on the conformational freedom of the structures. Notably, the angle was systematically lower in the simulation compared to the experimental measurements. A possible reason for the difference is that the experimental angles were measured from TEM images of surface deposited structures, which could have caused slight deformation of the structure due to the adherence to the TEM grid. In the 3D simulations we observe out-of-plane twisting in the structures (Figure B3e in appendix B). It is likely that deposition of structures on the surface compensates some of this out-of-plane twist by opening the structure slightly wider.

To sample the behaviour of the NE under tension, mimicking the action of T7RNAP, pulling and release simulations were performed by applying a constant force comparable to that exerted by the T7RNAP (16 pN) between the nucleotide covalently linked to the T7RNAP in experiments and the first nucleotide of the terminator sequence. In general, closing was faster than opening with rates of 0.28 ± 0.07 °/ns vs. 0.13 ± 0.06 °/ns for nNE and 0.28 ± 0.10 °/ns vs. 0.11 ± 0.05 °/ns for NE. These results are in accordance with what one would expect if we take into account that closing is a driven process while opening relies on releasing strain built up during the closing process and a component of Brownian motion. It also explains our experimental observation that nNEsoft showed a similar transcription rate as nNE (Figure B2j in appendix B). Interestingly, when secondary structures are prohibited the closing rate for nNE-NS was 0.29 ± 0.10 °/ns and the opening rate was 0.21 ± 0.07 °/ns. For NE-NS, the closing rate was 0.26 ± 0.08 °/ns and the opening rate was 0.22 ± 0.05 °/ns (Figure 20d). The similarity of the closing rates while doubling the opening rates when secondary structures are disallowed suggests that the secondary structures have little effect on the driven closing process, however they stabilize the closed state and impede the stochastic opening process. The transcribable sequence has a smaller impact on opening rates. In simulations where that region was deleted from the closed state and

allowed to re-open, the rate was slightly slower than the corresponding complete structure (0.08 ± 0.07 °/ns for NE_NB and 0.15 ± 0.06 °/ns for NE_NS_NB). No significant difference was observed between the rates of pulling and opening when NE (with the entirely transcribable dsDNA) and nNE (where the transcribable dsDNA contained two nicks) simulations initiated from the same starting configuration (Figure 20c, d). The simulation results hence suggest that the experimentally observed difference in closing and opening rates between NE and nNE can be attributed to supercoiling caused by the activity of the polymerase rather than to the mechanical properties conferred by the transcribable dsDNA to the NE itself.

The effect of temperature on the dynamic simulations was similar to the equilibrium simulations. The rates of opening and closing were slightly increased and the variance of replicates of closing and opening simulations increased at 37 °C, as would be expected due to the increase in the accessible phase space (Figure B3f,g in appendix B).

To get a better insight into the observed phenomenon of a reduced transcription rate when the transcribable dsDNA sequence is anchored only next to the polymerase we extracted the amount of time the promoter region spends next to the polymerase in the case when the dsDNA strand is fully anchored at both ends to the origami and when it is only attached next to the polymerase. As the T7RNAP is not explicitly represented in the simulations, an approximation of accessibility was made based on distance. The overall geometry of T7RNAP is roughly that of a sphere 10.35 nm in diameter²⁴³. Thus, to a first order approximation, transcription can only be initiated when the distance between the polymerase attachment site and promoter sequence are within 10.35 nm proximity (this approximation ignores the specific orientation required to initiate transcription). The simulations showed that in the nNE the promoter region spends $54.9 \pm 0.01\%$ of the time in the 10.35 nm radius but this time is reduced to $45.6 \pm 0.01\%$ when the dsDNA strand is only anchored next to the polymerase indicating how the higher number of degrees of freedom in the latter case reduces the chances of the promoter region to be in a favourable position to obtain efficient transcription. (Figure B3h in appendix B).

4.4 Driver-Follower Experiments

An important task of any engine that actively performs work is the ability to be coupled to passive moving parts for transmission of force or motion. Nature has found vast solutions for the transmission of force, primarily by myocytes or adherent cells, but examples demonstrating nanomechanical force and motion transmission by synthetic machines are scarce and the motion occurs mostly stochastic^{39,244-246}. To demonstrate that nNE

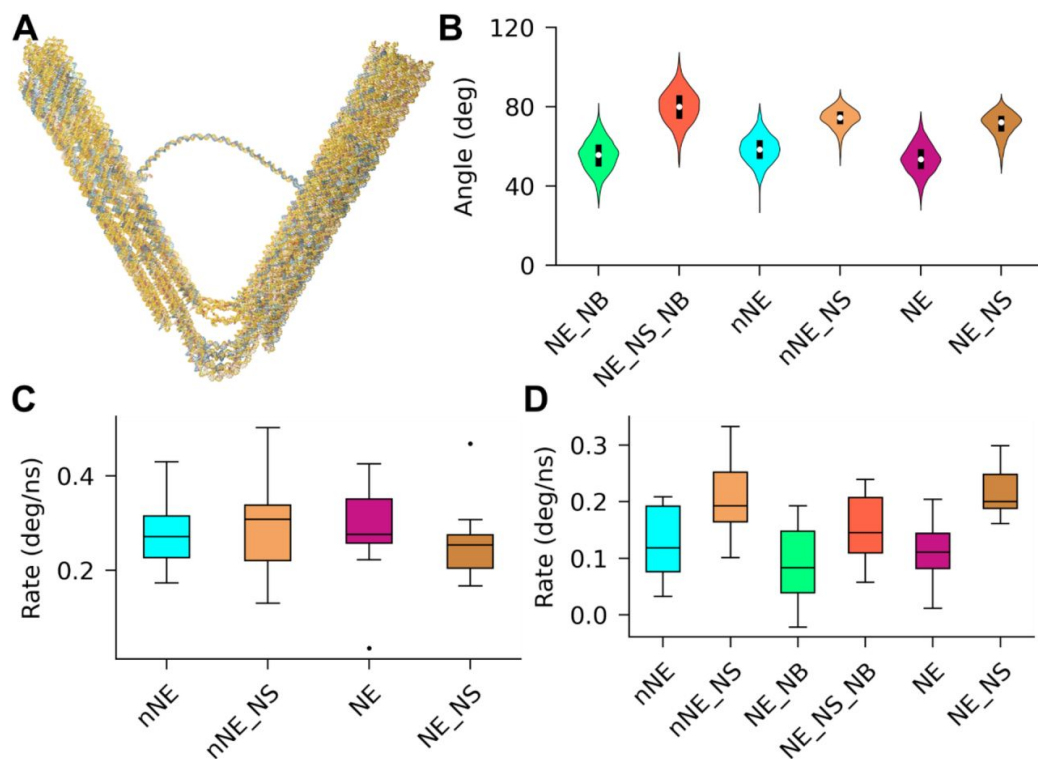


Figure 20. **Coarse-grained simulation of NE designs.** oxDNA simulations were performed to determine dynamic structural properties of the NE. (a) Mean structure of an equilibrium sampling simulation of nNE represented in oxDNA, (b) Equilibrium angle distribution of six designs during oxDNA simulation. NE - nano engine; nNE - nicked nano engine; NS - no structure, where base pairing was turned off for single stranded regions of the flexure to isolate the effect of secondary structures forming in the flexure; NB - no transcribable dsDNA bridge; here the transcribable dsDNA bridge was deleted using oxView's editing tools, to determine its effect on equilibrium and opening simulations, (c) Simulation-determined closing rates from pulling simulations, (d) simulation-determined opening rates from relaxation simulations where the forces from the pulling simulations are released and the structure is allowed to open again.

can act as non-stochastic, autonomous mechanical 'driver' (D, Figure 21a) and actively transmit its force to a passive part that follows its motion, we coupled nNE to a passive 'follower' (F, Figure 21b). We added five unique ssDNA overhangs protruding axially from the extremities of each origami arm of D and F. Each of the ten sequences in D allow for the hybridization of only the complementary 10 sequences on the F unit to form the rhomboid-shaped hetero dimer D-F (Fig 21c-e); formation of homodimers D-D or F-F is not possible due to the non complementarity of the sequences (Figure B4a,b in appendix B). To strengthen the connection between D and F, three LNA modifications²⁴⁷ were included into two of the overhanging sequences on each D-arm that connect to F (Figure B4b in appendix B, red).

As before for nNE, a large set of TEM images of only the rhomboidal D-F dimers (Figure B4c in appendix B) with and without transcription was analysed to obtain angle distributions under both conditions (Fig 21f). Box

plots of the distribution (Fig 21g) show that the median (Q2) shifts from 65° in absence of transcription (blue) to 52° in presence (orange). The quartile distribution Q1 shifts from 54° (no transcription, blue) to 39° (transcription, orange), and Q3 shifts from 75° (no transcription, blue) to 63° (transcription, orange). Moreover, the formation of the D-F complex apparently leads to slightly narrower, less skewed, and more symmetric angle distributions under both conditions (Figure B4d in appendix B, blue, orange) as compared to NE (red, green) suggesting that the connection of D and F has a stabilizing effect on the angle. These data demonstrate that transcribing D-F structures exhibit more acute angles on average than non-transcribing ones and indicate that the D-bound F unit follows the motion imposed by D.

This trend becomes even more apparent in case of a D-F complex in which the F unit has a completely single stranded flexure region (F-ss-hinge, Figure B4f,g in appendix B). In the 'ss-hinge'-design the interaction of dsDNA flat spring and ssDNA tension sequences is absent. Consequently, F does not assume a defined angle, which is reflected by a broad angle distribution median of 113° (Figure B4h,i in appendix B, olive, F-ss-hinge). However, when bound to D, the median shifts towards a more acute median angle of 71° (Figure B4h,i in appendix B, wine, D-F-ss-hinge no transcription) under no transcription, and a median of 56° under transcription conditions (Figure B4i in appendix B, green, D-F-ss-hinge transcription), comparable with the medians obtained for NE. The boxplots also show that the angle distributions become considerably narrower in the D-F-ss-hinge complex compared to F-ss-hinge: under no transcription the difference between Q3 and Q1 in D-F-ss-hinge measures only 23° while in the F-ss-hinge it measures 54° (Figure B4i in appendix B).

When the D unit was coupled to an F-derivative with a soft hinge, (F-soft-hinge; Figure B4j in appendix B) in which two staples are removed from the dsDNA flexure region, the transcription shows a behaviour that is comparable with the D-F sample. The median angle of the distribution shifts from 62° in absence of transcription (yellow, Figure B4k,l in appendix B) to 50° under transcription conditions (purple).

We next measured the transcription speed of different driver follower constructs relative to the single D (or nNE) unit. D was combined with the F-ss-hinge, F-soft-hinge, or the F unit (Fig 21h). Although the differences are small, we observed a significant increase of transcription rate for the D-F-ss-hinge and the D-F-soft-hinge, respectively, of 1.2 ± 0.2 ($p = 0.001$) and 1.2 ± 0.2 (p value 0.01), respectively, compared to nNE. The combination of D with F (D-F) showed no significant increase in transcription speed compared to nNE (Fig 21h, nNE vs. D-F). These results indicate that the combination of the active driver with a passive follower unit influences the closing and opening speed of the dimeric system. In case of the D-F we hypothesize that the addition of a complete hinge to the dimeric complex adds further resistance to the closing and this effect is compensated just enough by the reopening spring effect to result in no significant effect on the overall

transcription speed. In the case of the F-ss-hinge and the F-soft-hinge, however, the closing speed is not significantly affected due to less resistant springs but the reopening is enhanced by the higher entropic degree of spring areas resulting in a slightly increased transcription rate.

We simulated the D-F connected structure during the opening and closing process. When plotting the angle of the F unit against the corresponding angle in the D unit it is possible to notice that the angles have a linear correlation and are located along the bisector of the first quadrant ($y = x$). Deviation from the ideal line that divides the first quadrant with 45° can be attributed to inherent flexibility in the origami itself and more importantly in the contact points of the origami itself. Nevertheless, simulations show how, the angle that is assumed by the driver unit is transferred over to the follower unit in a linear one to one fashion (Figure 21i).

4.5 Conclusion

We describe the bottom-up construction of a biohybrid DNA-origami-based nanomachine that performs tasks that are fundamental for every contrivance requiring automated motion: An autonomous, fuel-driven, rhythmically pulsing DNA-nanoengine that can be easily latched to any type of passive DNA-origami-based 'follower' entities to which it then transmits its motion and force, thus constituting a genuine driver-follower pair.

Achieving autonomous active motion in a DNA-nanostructure constitutes a success in itself, but the demonstration that nNE can act as a driver of other devices adds considerable significance. Since DNA origami technology permits the bottom-up construction of robust structures with different mechanical properties that can span from mechanically rigid to bendable and compliant structures^{228,236,237,241}, all bearing the option of being combined into a single architecture, the versatility of mechanical power transmission by nNE is vast. Although the different passive follower structures used here are fairly simple, the prototypical design of D-F suggests that nNE should be applicable in other DNA-nanostructures to achieve considerably larger structural rearrangements as exemplified before in non-autonomous systems⁹⁹. nNE provides an actively moving nano engine that operates fully autonomously; once the NTP-fuel is added to the system, the structure sets itself in motion, and continues pulsation over several hours without further input. It can even be envisioned introducing a switchable clutch mechanism that straightforwardly allows detaching the driver from a coupled follower and attaching it to a different one while the engine is still running. For example by controlling the

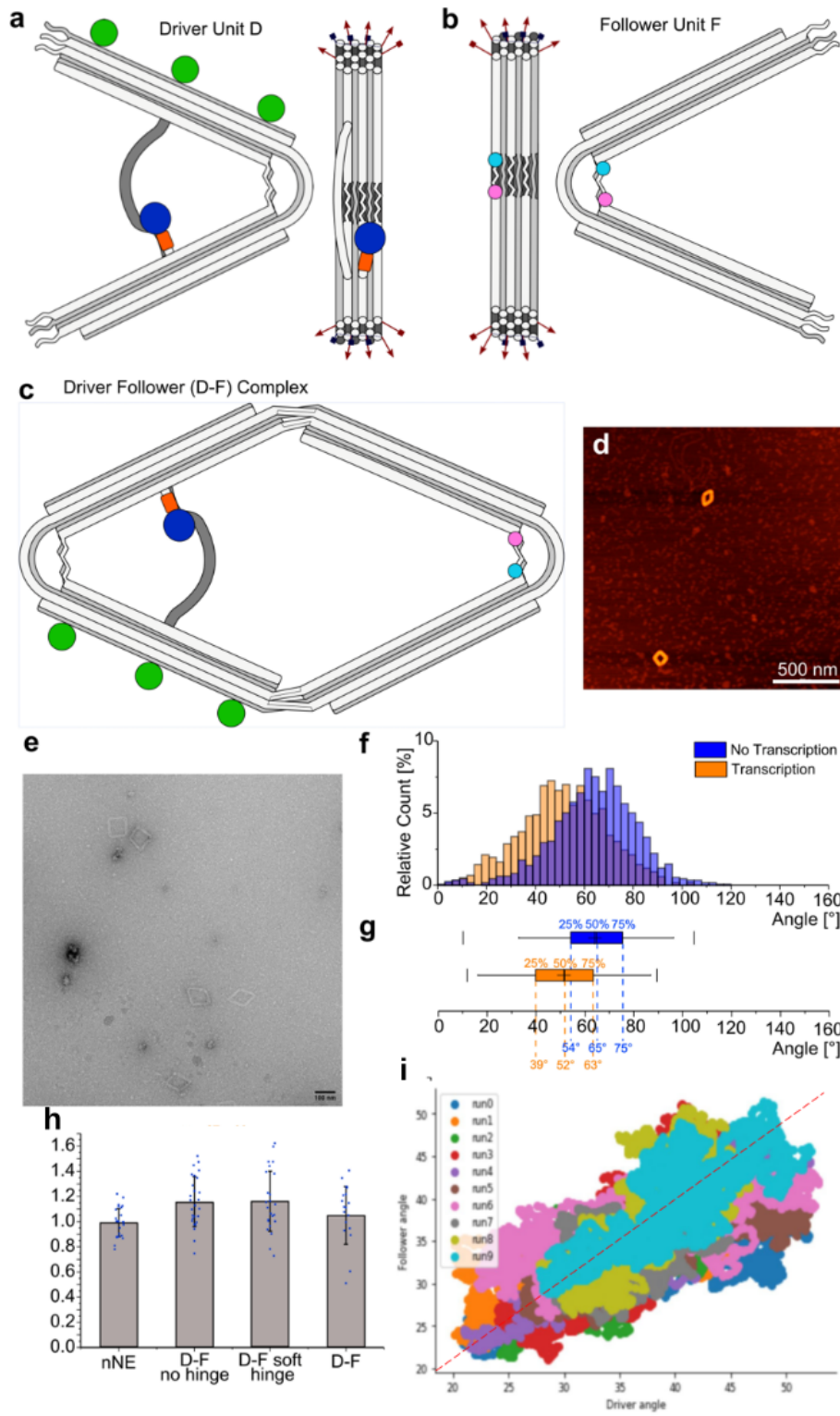


Figure 21. (Previous page.) **nNE drives a passive follower unit.** (a) Schematic representation of the nNE driver D and the passive follower unit F in (b). The structures are modified to have ssDNA overhangs protruding from the stiff origami arms. Five ssDNA overhangs (shown as red arrows) were introduced on each arm for a total of 10 unique overhangs on each origami. The sequences are designed to allow the connection of only the active structure to the passive structure. Biotin modifications are introduced only on the driver unit and are shown as green dots. A Cy3 and Cy5 FRET pair in F are shown with a pink and cyan dot, (c) the assembly of the D–F heterodimer is achieved by equimolarly combining the two parts of the system and allowing a thermal annealing overnight. After 4.5 h of transcription experiments at 37 °C the integrity of the system is confirmed by AFM (d) size bar: 500 nm, and TEM (e) size bar: 100 nm, (f) Analysis of the angle distribution from TEM images of the D–F dimer complex for samples that did not (n=1074, blue) and did undergo transcription (n=1190, orange). The distribution was plotted as bar graphs with a bin with of 3°, the count of each bar was divided by the total number of counts, and displayed as the relative count percentages. The distribution significantly shifts towards more acute angles under transcription ($p = 8 \times 10^{-64}$), indicating the formation of structures with smaller angles during transcription; (g) Boxplot representation of distributions. Blue: boxplot of D–F without transcription, orange: box plot of D–F with transcription. The boxes represent the distribution at first quartile (Q1; 25%), second quartile or median (Q2; 50%) and the third quartile (Q3; 75%) with dashed lines indicating the angular value in the corresponding color. The whiskers represent 1.5-times the box size while the thin vertical lines represent the 1% and the 99% percentile of the distribution. The thin cross indicates the average angle of the distribution. Average angles of distribution: D–F in absence of transcription $64^\circ \pm 17^\circ$ (S.D. error), D–F with transcription $51^\circ \pm 17^\circ$, (h) The transcription speed of different D–F complexes was measured and compared to the transcription speed of D (nNE) alone (n = 27). Transcription speeds: D–F–ss–hinge 1.2 ± 0.2 (n = 27, p = 0.001), D–F–soft–hinge 1.2 ± 0.2 (n = 18, p = 0.01), D–F complex 1.1 ± 0.2 (n = 24; p = 0.06). Bargraph: average value with error S.D. (i) Simulation traces showing D–F coupling during 10 pulling simulations of the D–F complex. Dashed line at $y = x$ for reference.

hybridisation of the D–F connecting ODNs with light-switchable isomers, as has been shown for several DNA nanomachines^{227,248–252}.

By expanding the complexity of the structural features that can be easily attached to the origami by single stranded overhangs or other chemical or biochemical modification, there is potentially no limit to the range of motion or the type of motion that can be linear, rotative, pulsatile or contractile, that could be achieved by introducing the nano engine as an active core component at the heart of bigger passive nano machines, the function of which depends on downstream sequences of motion.

4.6 Author Contributions

M. F., Mat. C., and J. V. developed the here described concept of the T7RNAP–driven nanoengine. Mat.C. performed and designed, with M. F. and initially also with J. V., most of the included studies. M. F. supervised the research project, assisted by J. V. Mar. C. performed all TEM studies. E. P. and P. Š. carried out all molecular modelling. All authors discussed the experimental results and contributed to writing the manuscript, with Mat. C. and M. F. doing the bulk of the writing.

4.7 Additional Information

The authors declare no competing financial interests. The input files and analysis scripts for the simulation portions of this work are available at <https://github.com/sulcgroup/hinges>, while the full oxDNA trajectories can be found at

https://drive.google.com/drive/folders/1KoocIZRPcRJ7us0q695_ya3cXNxYUn2t. Supplementary Datasets S1-S3 can be found at

<https://drive.google.com/drive/folders/1gp6IGHHh55W4BCZR7tnTNTAuUsph-GQQ?usp=sharing>.

4.8 Acknowledgements

We thank V. Fieberg, and D. Keppner for technical assistance, and Werner Kühlbrandt, MPI of Biophysics, Frankfurt, for providing access to electron microscopy facilities. This work was supported by the Alexander von Humboldt Foundation, the Max-Planck Society and the University of Bonn. P. Š. and E. P. acknowledge the use of the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number TG-BIO210009.

CONCLUSION AND FUTURE OUTLOOKS

Scientific software is notorious for its poor user design and rapid depreciation. In order for a scientific field to reach maturity, it needs to develop a consistent set of tools which not only facilitate researchers in achieving their goals, but also provides the continuity to allow institutional knowledge to pass through successive generations of trainees and allow comparisons with past realizations. Nucleic acid nanotechnology remains a field of high flux, with many new tools created every year to support nanoscale engineers in designing and modeling their vision. Through the past decade, however, the oxDNA/RNA model and CaDNAo have remained consistently popular despite attempts to create newer tools. This beginning of crystallized software choices has led to a need for a new class of tools: software that supports, expands and connects legacy tools together. The work presented at this dissertation is a first step towards creating a robust integrated DNA/RNA design and simulation environment.

Chapters 2 and 3, introduced two new pieces of software which themselves do not answer novel scientific questions, but rather make the oxDNA simulation engine more accessible and the data it produces easier to analyze. These tools support and expand existing infrastructure for working with the oxDNA/RNA models (Figure 22). Through the improved simulation setup and results visualization options provided by oxView, it becomes more possible to use oxDNA as an iterative design tool where oxDNA simulations are used to identify weaknesses in designs and oxView is used to perform fine-tune edits and test different designs. The structure studied in chapter 4 is a fantastic example of this. Experimental characterization of the nanoengine (NE) left many questions unanswered due to the limited resolution of microscopy techniques and the effects of the surface. Combined with the format conversion tools introduced in¹⁷⁴, oxView was used to take the structure from the flat structure produced by conversion from CaDNAo to the many variations on the structure shown in chapter 4 (Figure 23). Thanks to these simulations, I was able to identify a deficiency in the design in that the scaffold strand was forming hairpins in a region that was intended to be single stranded. These hairpin structures had a significant impact on the overall behavior of the NE, resulting in a smaller equilibrium opening angle and slower re-opening kinetics than would have been observed for a design which used a custom, orthogonal scaffold²⁵³.

This pattern of design-convert-build-simulate-analyze-edit has become integral in the design of complex

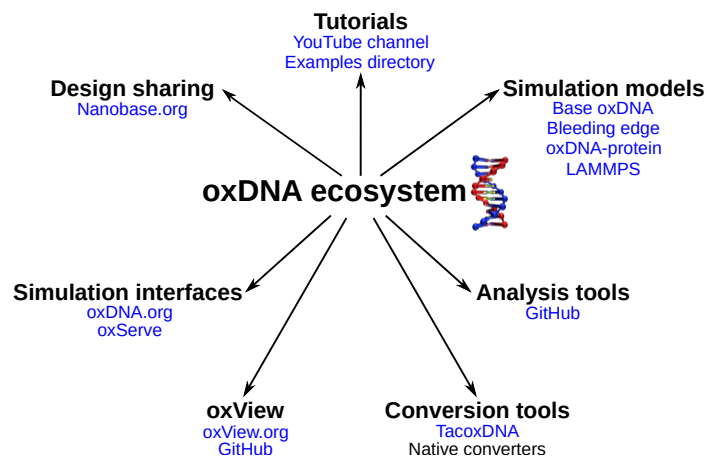


Figure 22. **The oxDNA ecosystem** The tools presented in this dissertation are part of a larger ecosystem of tools which implement, expand and support the oxDNA/RNA model.

nanostructures. Three of those four steps using the tools demonstrated in chapter 2. There have been numerous papers in recent years^{54,254–256} which use this pattern to demonstrate some of the largest and most complex DNA-based structures ever developed. It has been an honor to see the tools that I led the development of become so integral to the field and play a key role in the maturation of DNA nanotechnology into a field that is beginning to have applications beyond simply making interesting structures at the nanoscale.

Both oxView and oxDNA Analysis Tools are living pieces of software; because of their core role in the research performed by the Šulc lab and others, myself and the other developers have continued to develop and maintain the tools. Due to optimizations to the file readers and parallelization functions, oxDNA Analysis Tools now runs computations 10 – 100x faster than it did when chapter 2 was published and the oxView user interface has been completely redesigned with many more editing features as well as the ability to prepare simplified protein models¹⁶¹. OxDNA.org has run more than 10,000 jobs for hundreds of users since its original publication. At the time of writing, it has only been a year since the original publication, however the webserver has already proven useful to many projects^{106,257–261}. It has introduced many previously experimental-only labs to the potential of using molecular simulation to further explore their designs.

Looking to the future, there's still a lot to improve in the nanotechnology software environment. As we move towards more complex and functional nanomaterials, the diversity of the chemistry has continuously grown. DNA nanostructures are no longer just DNA, but are now decorated with proteins²⁶², metal nanoparticles²⁶³, and hydrophobic moieties²⁶⁴ in order to perform functional tasks, control light, and interface with cells and cell-like vesicles. Currently, no tool offers a robust method to design with these other molecule types the way CaDNAno lets us design with DNA. Towards that end, I have been involved in discussions with

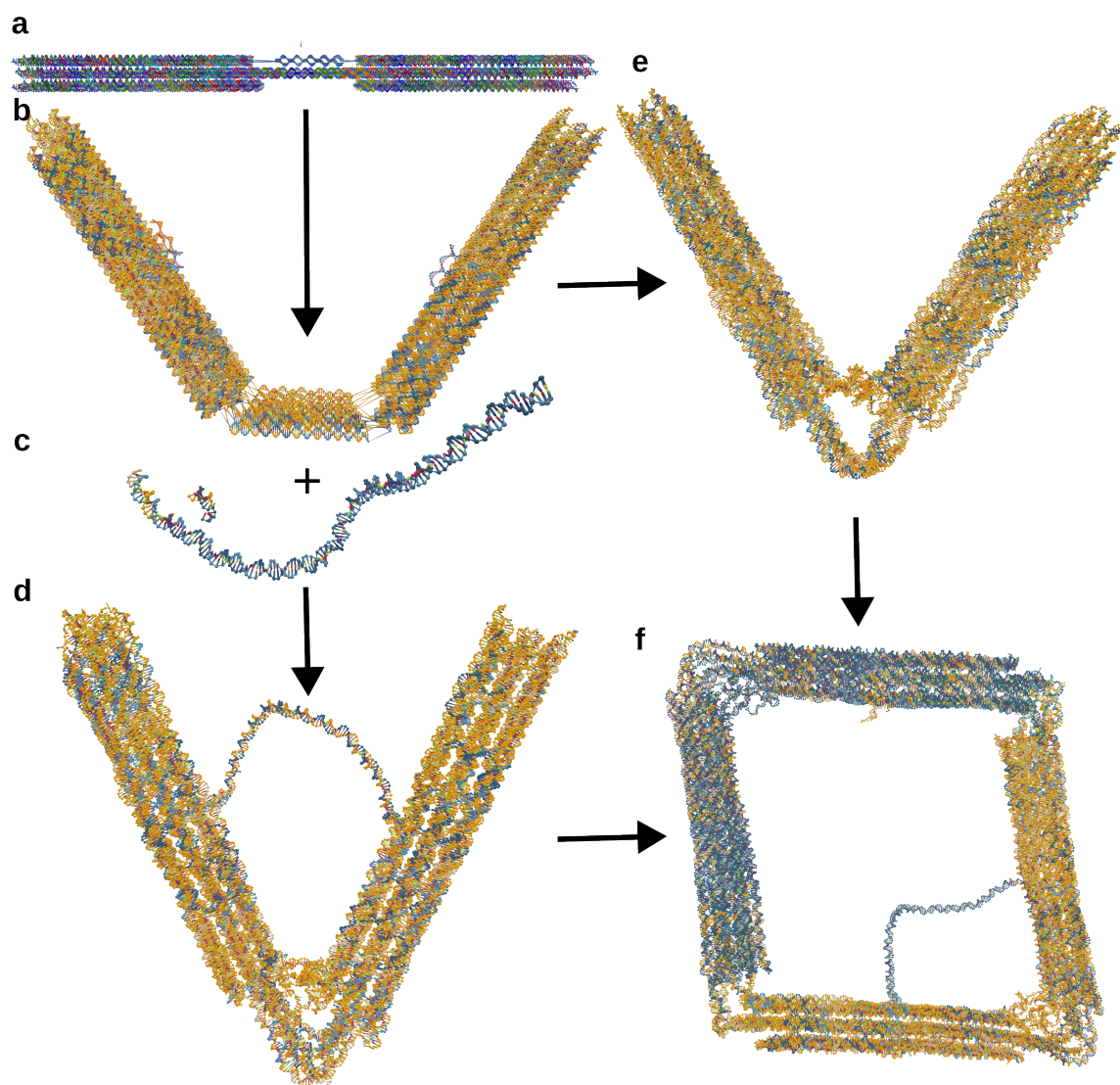


Figure 23. **Using oxView to build the leaf-spring nanoengine (NE).** (a) The design as exported from CaDNAo. In CaDNAo all helices are parallel and creating off-lattice components is difficult. (b) OxView's translation, rotation, and rigid-body-dynamics tools were used to get the NE into roughly the correct 3D geometry. (c) The polymerase attachment site and the dsDNA bridge were built using oxView's freeform design tools. (d) The relaxed NE after ligation of the freeform components and initial relaxation (e) The no-bridge NE (NB_NE) was relaxed straight from (b). (f) The driver-follower structure was created by building the connections (see Figure B2 in oxView and pulling the structure together with artificial forces.

labs around the world as to the kinds of features we would like to see in future nano-CAD software packages. One of the first products of these conversations is a draft of a highly-verbose file format, the Universal Nanotechnology Format²⁶⁵ which would allow software tools to both work with familiar formats — the CaDNAno grid of virtual helices and the positions and rotations of the oxDNA format as well as other molecule types and representations including PDB files for fully-atomistic protein structures. This draft format is far from perfect, but we hope it will present the opportunity for future software developers to think about how we interact with data structures and help future tools work with additional materials.

Another development that is needed to improve the ability of the nanotechnology field to build devices from diverse materials are the tools to model them. The oxDNA/RNA model has been a fantastic success for DNA and RNA, however is limited to those contexts plus a very simplified representation of proteins. Furthermore, as structures grow larger and more complex, the nucleotide resolution used by oxDNA/RNA may be too high-resolution to characterize structures at the timescales required. In light of this, the future of the modeling field needs to be to move towards more multi-scale and multi-material models. A model which could capture less-relevant or stiffer parts of a structure as rigid, convex polyhedra while representing components of interest at the nucleotide (or subnucleotide) scale would greatly speed up, and improve the accuracy of molecular models for nucleic acid nanotechnology. Multi-material models, including DNA, RNA, proteins, small molecules, lipids, and metal nanoparticles is also a goal the field should strive for. The MARTINI²⁶⁶ force field is an attempt at a model of this type, however this type of model still needs significant work; MARTINI is notoriously bad at correctly reproducing the mechanics of DNA helices²⁶⁷.

While there is still much work to be done, the future of biological nanotechnology is bright. Between more complex, robust and functional DNA and RNA structures, recent successes in *de novo* protein design and efforts to put it all together in fully synthetic cell-like micromachines, there are beautiful and powerful creations being developed all the time. Perhaps soon we will live in a world where multiplexed diagnostic tools built from DNA computers, precision RNAi medicines are delivered with DNA nanocarriers, and fine chemicals are produced on DNA-structured catalysis networks inside compartmentalized vesicles are all realities and not the stuff of science fiction and marquee fraud cases. Whatever developments the future holds, it's going to happen on the back of good software which magnifies human creativity making the hard parts of design easy and the easy parts trivial.

REFERENCES

1. Peters, R. J. *et al.* Nanomaterials for products and application in agriculture, feed and food. *Trends in Food Science & Technology* **54**, 155–164 (2016).
2. Ezhilarasi, P., Karthik, P., Chhanwal, N. & Anandharamakrishnan, C. Nanoencapsulation techniques for food bioactive components: a review. *Food and Bioprocess Technology* **6**, 628–647 (2013).
3. Kato, T. *et al.* Enhanced performance of 50 nm Ultra-Narrow-Body silicon carbide MOSFETs based on FinFET effect in 2020 32nd International Symposium on Power Semiconductor Devices and ICs (ISPSD) (2020), 62–65.
4. Schoenmaker, L. *et al.* mRNA-lipid nanoparticle COVID-19 vaccines: Structure and stability. *International Journal of Pharmaceutics* **601**, 120586 (2021).
5. Brader, M. L. *et al.* Encapsulation state of messenger RNA inside lipid nanoparticles. *Biophysical journal* **120**, 2766–2770 (2021).
6. Veisi, H., Ghorbani, M. & Hemmati, S. Sonochemical in situ immobilization of Pd nanoparticles on green tea extract coated Fe₃O₄ nanoparticles: an efficient and magnetically recyclable nanocatalyst for synthesis of biphenyl compounds under ultrasound irradiations. *Materials Science and Engineering: C* **98**, 584–593 (2019).
7. Yuan, Y., Gao, Y., Zhao, J. & Mao, L. Characterization and stability evaluation of β -carotene nanoemulsions prepared by high pressure homogenization under various emulsifying conditions. *Food Research International* **41**, 61–68 (2008).
8. Feynman, R. P. There's plenty of room at the bottom. *Journal of microelectromechanical systems* **1**, 60–66 (1992).
9. Maragò, O. M., Jones, P. H., Gucciardi, P. G., Volpe, G. & Ferrari, A. C. Optical trapping and manipulation of nanostructures. *Nature nanotechnology* **8**, 807–819 (2013).
10. Hansen, L. T., Kühle, A., Sørensen, A. H., Bohr, J. & Lindelof, P. A technique for positioning nanoparticles using an atomic force microscope. *Nanotechnology* **9**, 337 (1998).
11. Chen, Y. Nanofabrication by electron beam lithography and its applications: A review. *Microelectronic Engineering* **135**, 57–72 (2015).
12. Nellis, S. TSMC says has begun construction at its Arizona chip factory site. *Reuters*. www.reuters.com/technology/tsmc-says-construction-has-started-arizona-chip-factory-2021-06-01/ (2021).
13. Stokes, G. G. *et al.* On the effect of the internal friction of fluids on the motion of pendulums (1851).
14. Purcell, E. M. Life at low Reynolds number. *American journal of physics* **45**, 3–11 (1977).
15. Brown, R. XXVII. A brief account of microscopical observations made in the months of June, July and August 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies. *The philosophical magazine* **4**, 161–173 (1828).
16. Busing, W. R. & Levy, H. A. The effect of thermal motion on the estimation of bond lengths from diffraction measurements. *Acta Crystallographica* **17**, 142–146 (1964).

17. Lee, S. H. & Kapral, R. Mesoscopic description of solvent effects on polymer dynamics. *The Journal of chemical physics* **124**, 214901 (2006).
18. Vikesland, P. J., Heathcock, A. M., Rebodos, R. L. & Makus, K. E. Particle size and aggregation effects on magnetite reactivity toward carbon tetrachloride. *Environmental science & technology* **41**, 5277–5283 (2007).
19. Ohmann, A. *et al.* Controlling aggregation of cholesterol-modified DNA nanostructures. *Nucleic acids research* **47**, 11441–11451 (2019).
20. Shrestha, S., Wang, B. & Dutta, P. Nanoparticle processing: Understanding and controlling aggregation. *Advances in Colloid and Interface Science* **279**, 102162 (2020).
21. Bishop, B. *Nanoengineer-1* 2014. <https://github.com/kanzure/nanoengineer>.
22. OneAngstrom. 2022. <https://www.samson-connect.net/>.
23. Yu, H., Peng, Y., Yang, Y. & Li, Z.-Y. Plasmon-enhanced light–matter interactions and applications. *npj Computational Materials* **5**, 1–14 (2019).
24. Cai, W. & Shalaev, V. M. *Optical metamaterials* **6011** (Springer, 2010).
25. Zasedatelev, A. V. *et al.* Single-photon nonlinearity at room temperature. *Nature* **597**, 493–497 (2021).
26. Willets, K. A. & Van Duyne, R. P. Localized surface plasmon resonance spectroscopy and sensing. *Annual review of physical chemistry* **58**, 267–297 (2007).
27. Kaiser, S. K. *et al.* Performance descriptors of nanostructured metal catalysts for acetylene hydrochlorination. *Nature Nanotechnology*, 1–7 (2022).
28. Berg, O. G., Winter, R. B. & Von Hippel, P. H. Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry* **20**, 6929–6948 (1981).
29. Schrodinger, E. *What is life? The physical aspect of the living cell* (Cambridge University Press, 1944).
30. Girelli, G. *et al.* GPSeq reveals the radial organization of chromatin in the cell nucleus. *Nature biotechnology* **38**, 1184–1193 (2020).
31. Ramakrishnan, V. Ribosome structure and the mechanism of translation. *Cell* **108**, 557–572 (2002).
32. Houdusse, A. & Sweeney, H. L. How myosin generates force on actin filaments. *Trends in biochemical sciences* **41**, 989–997 (2016).
33. Nurk, S. *et al.* The complete sequence of a human genome. *Science* **376**, 44–53 (2022).
34. Watson, J. D. & Crick, F. H. Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature* **171**, 737–738 (1953).
35. Seeman, N. C. Nucleic acid junctions and lattices. *Journal of theoretical biology* **99**, 237–247 (1982).
36. Holliday, R. A mechanism for gene conversion in fungi. *Genetics Research* **5**, 282–304 (1964).
37. Praetorius, F. *et al.* Biotechnological mass production of DNA origami. *Nature* **552**, 84–87 (2017).

38. Kallenbach, N. R., Ma, R.-I. & Seeman, N. C. An immobile nucleic acid junction constructed from oligonucleotides. *Nature* **305**, 829–831 (1983).
39. Chen, J. & Seeman, N. C. Synthesis from DNA of a molecule with the connectivity of a cube. *Nature* **350**, 631–633 (1991).
40. Zhang, Y. & Seeman, N. C. Construction of a DNA-truncated octahedron. *Journal of the American Chemical Society* **116**, 1661–1669 (1994).
41. Goodman, R. P., Berry, R. M. & Turberfield, A. J. The single-step synthesis of a DNA tetrahedron. *Chemical Communications*, 1372–1373 (2004).
42. Winfree, E., Liu, F., Wenzler, L. A. & Seeman, N. C. Design and self-assembly of two-dimensional DNA crystals. *Nature* **394**, 539–544 (1998).
43. Yan, H., Park, S. H., Finkelstein, G., Reif, J. H. & LaBean, T. H. DNA-templated self-assembly of protein arrays and highly conductive nanowires. *science* **301**, 1882–1884 (2003).
44. Winfree, E. *Algorithmic self-assembly of DNA* (California Institute of Technology, 1998).
45. Mao, C., LaBean, T. H., Reif, J. H. & Seeman, N. C. Logical computation using algorithmic self-assembly of DNA triple-crossover molecules. *Nature* **407**, 493–496 (2000).
46. Rothmund, P. W. K., Papadakis, N., Winfree, E. & Condon, A. Algorithmic self-assembly of DNA Sierpinski triangles. *PLoS biology* **2**, e424 (2004).
47. Adleman, L. M. Molecular computation of solutions to combinatorial problems. *science* **266**, 1021–1024 (1994).
48. Liu, W., Zhong, H., Wang, R. & Seeman, N. C. Crystalline two-dimensional DNA-origami arrays. *Angewandte Chemie International Edition* **50**, 264–267 (2011).
49. Tikhomirov, G., Petersen, P. & Qian, L. Programmable disorder in random DNA tilings. *Nature nanotechnology* **12**, 251–259 (2017).
50. Douglas, S. M. *et al.* Self-assembly of DNA into nanoscale three-dimensional shapes. *Nature* **459**, 414–418 (2009).
51. Dietz, H., Douglas, S. M. & Shih, W. M. Folding DNA into twisted and curved nanoscale shapes. *Science* **325**, 725–730 (2009).
52. Han, D. *et al.* DNA origami with complex curvatures in three-dimensional space. *Science* **332**, 342–346 (2011).
53. Tikhomirov, G., Petersen, P. & Qian, L. Fractal assembly of micrometre-scale DNA origami arrays with arbitrary patterns. *Nature* **552**, 67–71 (2017).
54. Yao, G. *et al.* Meta-DNA structures. *Nature chemistry* **12**, 1067–1075 (2020).
55. Wintersinger, C. M. *et al.* Multi-micron crisscross structures from combinatorially assembled DNA-origami slats. *bioRxiv* (2022).
56. Douglas, S. M., Bachelet, I. & Church, G. M. A logic-gated nanorobot for targeted transport of molecular payloads. *Science* **335**, 831–834 (2012).

57. Jiang, S., Ge, Z., Mou, S., Yan, H. & Fan, C. Designer DNA nanostructures for therapeutics. *Chem* **7**, 1156–1179 (2021).
58. Chang, M., Yang, C.-S. & Huang, D.-M. Aptamer-conjugated DNA icosahedral nanoparticles as a carrier of doxorubicin for cancer therapy. *ACS nano* **5**, 6156–6163 (2011).
59. Li, S. *et al.* A DNA nanorobot functions as a cancer therapeutic in response to a molecular trigger in vivo. *Nature biotechnology* **36**, 258–264 (2018).
60. Wu, T. *et al.* A nanobody-conjugated DNA nanoplatform for targeted platinum-drug delivery. *Angewandte Chemie* **131**, 14362–14366 (2019).
61. Li, J. *et al.* Self-assembled multivalent DNA nanostructures for noninvasive intracellular delivery of immunostimulatory CpG oligonucleotides. *ACS nano* **5**, 8783–8789 (2011).
62. Schuller, V. J. *et al.* Cellular immunostimulation by CpG-sequence-coated DNA origami structures. *ACS nano* **5**, 9696–9702 (2011).
63. Wang, C., Sun, W., Wright, G., Wang, A. Z. & Gu, Z. Inflammation-triggered cancer immunotherapy by programmed delivery of CpG and anti-PD1 antibody. *Advanced Materials* **28**, 8912–8920 (2016).
64. Keum, J.-W., Ahn, J.-H. & Bermudez, H. Design, assembly, and activity of antisense DNA nanostructures. *Small* **7**, 3529–3535 (2011).
65. Roh, Y. H. *et al.* Layer-by-layer assembled antisense DNA microsponge particles for efficient delivery of cancer therapeutics. *ACS nano* **8**, 9767–9780 (2014).
66. Li, J. *et al.* Self-assembly of DNA nanohydrogels with controllable size and stimuli-responsive property for targeted gene regulation therapy. *Journal of the American Chemical Society* **137**, 1412–1415 (2015).
67. Sun, W. *et al.* Precise pitch-scaling of carbon nanotube arrays within three-dimensional DNA nanotrenches. *Science* **368**, 874–877 (2020).
68. Zhao, M. *et al.* DNA-directed nanofabrication of high-performance carbon nanotube field-effect transistors. *Science* **368**, 878–881 (2020).
69. Iwaki, M., Wickham, S., Ikezaki, K., Yanagida, T. & Shih, W. A programmable DNA origami nanospring that reveals force-induced adjacent binding of myosin VI heads. *Nature communications* **7**, 1–10 (2016).
70. Funke, J. J. *et al.* Uncovering the forces between nucleosomes using DNA origami. *Science advances* **2**, e1600974 (2016).
71. Zhang, Y., Ge, C., Zhu, C. & Salaita, K. DNA-based digital tension probes reveal integrin forces during early cell adhesion. *Nature communications* **5**, 1–10 (2014).
72. Endo, M., Katsuda, Y., Hidaka, K. & Sugiyama, H. Regulation of DNA methylation using different tensions of double strands constructed in a defined DNA nanostructure. *Journal of the American Chemical Society* **132**, 1592–1597 (2010).
73. Pfitzner, E. *et al.* Rigid DNA beams for high-resolution single-molecule mechanics. *Angewandte Chemie International Edition* **52**, 7766–7771 (2013).

74. Nickels, P. C. *et al.* Molecular force spectroscopy with a DNA origami-based nanoscopic force clamp. *Science* **354**, 305–307 (2016).
75. Shaw, A. *et al.* Binding to nanopatterned antigens is dominated by the spatial tolerance of antibodies. *Nature nanotechnology* **14**, 184–190 (2019).
76. Veneziano, R. *et al.* Role of nanoscale antigen organization on B-cell activation probed using DNA origami. *Nature nanotechnology* **15**, 716–723 (2020).
77. Woods, D. *et al.* Diverse and robust molecular algorithms using reprogrammable DNA self-assembly. *Nature* **567**, 366–372 (2019).
78. Rothmund, P. W. Folding DNA to create nanoscale shapes and patterns. *Nature* **440**, 297–302 (2006).
79. Sigl, C. *et al.* Programmable icosahedral shell system for virus trapping. *Nature materials* **20**, 1281–1289 (2021).
80. Broyde, S. & Hingerty, B. 'A'forms of RNAs in single strands, duplexes and RNA-DNA hybrids. *Nucleic acids research* **5**, 2729–2742 (1978).
81. Polacek, N. & Mankin, A. S. The ribosomal peptidyl transferase center: structure, function, evolution, inhibition. *Critical reviews in biochemistry and molecular biology* **40**, 285–311 (2005).
82. Kruger, K. *et al.* Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *cell* **31**, 147–157 (1982).
83. Hausner, G., Hafez, M. & Edgell, D. R. Bacterial group I introns: mobile RNA catalysts. *Mobile DNA* **5**, 1–12 (2014).
84. AbouHaidar, M. G. & Ivanov, I. G. Non-enzymatic RNA hydrolysis promoted by the combined catalytic activity of buffers and magnesium ions. *Zeitschrift für Naturforschung C* **54**, 542–548 (1999).
85. Beckert, B. & Masquida, B. in *Rna* 29–41 (Springer, 2011).
86. Han, D. *et al.* Single-stranded DNA and RNA origami. *Science* **358**, eaao2648 (2017).
87. Geary, C., Grossi, G., McRae, E. K., Rothmund, P. W. & Andersen, E. S. RNA origami design tools enable cotranscriptional folding of kilobase-sized nanoscaffolds. *Nature chemistry* **13**, 549–558 (2021).
88. Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *nature* **391**, 806–811 (1998).
89. Guo, P. *et al.* Engineering RNA for targeted siRNA delivery and medical application. *Advanced drug delivery reviews* **62**, 650–666 (2010).
90. Majikes, J. M., Patrone, P. N., Kearsley, A. J., Zwolak, M. & Liddle, J. A. Failure Mechanisms in DNA Self-Assembly: Barriers to Single-Fold Yield. *ACS nano* **15**, 3284–3294 (2021).
91. Seeman, N. C. Interactive design and manipulation of macro-molecular architecture utilizing nucleic acid junctions. *Journal of Molecular Graphics* **3**, 34–39 (1985).
92. Seeman, N. C. De novo design of sequences for nucleic acid structural engineering. *Journal of biomolecular structure and dynamics* **8**, 573–581 (1990).

93. Birac, J. J., Sherman, W. B., Kopatsch, J., Constantinou, P. E. & Seeman, N. C. Architecture with GIDEON, a program for design in structural DNA nanotechnology. *Journal of Molecular Graphics and Modelling* **25**, 470–480 (2006).
94. Williams, S. *et al.* Tiamat: a three-dimensional editing tool for complex DNA structures in *International Workshop on DNA-Based Computers* (2008), 90–101.
95. Yang, Y. *et al.* An infectious virus-like particle built on a programmable icosahedral DNA framework. *ChemRxiv* (2022).
96. Dey, S. *et al.* A reversibly gated protein-transporting membrane channel made of DNA. *Nature Communications* **13**, 1–12 (2022).
97. Douglas, S. M. *et al.* Rapid prototyping of 3D DNA-origami shapes with caDNAno. *Nucleic acids research* **37**, 5001–5006 (2009).
98. Tian, Y. *et al.* Ordered three-dimensional nanomaterials using DNA-prescribed and valence-controlled material voxels. *Nature materials* **19**, 789–796 (2020).
99. Gerling, T., Wagenbauer, K. F., Neuner, A. M. & Dietz, H. Dynamic DNA devices and assemblies formed by shape-complementary, non-base pairing 3D components. *Science* **347**, 1446–1452 (2015).
100. Zhang, T. *et al.* 3D DNA origami crystals. *Advanced Materials* **30**, 1800273 (2018).
101. Glaser, M. *et al.* The art of designing DNA nanostructures with CAD software. *Molecules* **26**, 2287 (2021).
102. Doty, D., Lee, B. L. & Stérin, T. *scadnano: A Browser-Based, Scriptable Tool for Designing DNA Nanostructures* in *26th International Conference on DNA Computing and Molecular Programming (DNA 26)* (2020).
103. Nicolas, L. & Schabanel, N. *ENSnano: A 3D Modeling Software for DNA Nanostructures* in *27th International Conference on DNA Computing and Molecular Programming (DNA 27)* (2021).
104. Benson, E. *et al.* DNA rendering of polyhedral meshes at the nanoscale. *Nature* **523**, 441–444 (2015).
105. Benson, E. *et al.* Computer-Aided Production of Scaffolded DNA Nanostructures from Flat Sheet Meshes. *Angewandte Chemie* **128**, 9015–9018 (2016).
106. Jun, H. *et al.* Rapid prototyping of arbitrary 2D and 3D wireframe DNA origami. *Nucleic acids research* **49**, 10265–10274 (2021).
107. Veneziano, R. *et al.* Designer nanoscale DNA assemblies programmed from the top down. *Science* **352**, 1534–1534 (2016).
108. Jun, H. *et al.* Autonomously designed free-form 2D DNA origami. *Science advances* **5**, eaav0655 (2019).
109. Jun, H. *et al.* Automated sequence design of 3D polyhedral wireframe DNA origami with honeycomb edges. *ACS nano* **13**, 2083–2093 (2019).
110. Jun, H., Wang, X., Bricker, W. P. & Bathe, M. Automated sequence design of 2D wireframe DNA origami with honeycomb edges. *Nature communications* **10**, 1–9 (2019).

111. Huang, C.-M., Kucinic, A., Johnson, J. A., Su, H.-J. & Castro, C. E. Integrated computer-aided engineering and design for DNA assemblies. *Nature Materials* **20**, 1264–1271 (2021).
112. De Llano, E. *et al.* Adenita: interactive 3D modelling and visualization of DNA nanostructures. *Nucleic acids research* **48**, 8269–8275 (2020).
113. Kut'ák, D. *et al.* CATANA: an online modelling environment for proteins and nucleic acid nanostructures. *Nucleic Acids Research* (2022).
114. Poppleton, E. *et al.* Design, optimization and analysis of large DNA and RNA nanostructures through interactive visualization, editing and molecular simulation. *Nucleic Acids Research* (May 2020).
115. Das, R. & Baker, D. Macromolecular modeling with rosetta. *Annual review of biochemistry* **77**, 363–382 (2008).
116. Han, D. *et al.* DNA gridiron nanostructures based on four-arm junctions. *Science* **339**, 1412–1415 (2013).
117. Tian, Y. *et al.* Prescribed nanoparticle cluster architectures and low-dimensional arrays built using octahedral DNA origami frames. *Nature nanotechnology* **10**, 637–644 (2015).
118. Bohlin, J. *et al.* Design and simulation of DNA, RNA and hybrid protein–nucleic acid nanostructures with oxView. *Nature Protocols*, 1–27 (2022).
119. Poppleton, E., Mallya, A., Dey, S., Joseph, J. & Šulc, P. Nanobase. org: a repository for DNA and RNA nanostructures. *Nucleic acids research* **50**, D246–D252 (2022).
120. Sugimoto, Y. *et al.* Chemical identification of individual surface atoms by atomic force microscopy. *Nature* **446**, 64–67 (2007).
121. Eigler, D. M. & Schweizer, E. K. Positioning single atoms with a scanning tunnelling microscope. *Nature* **344**, 524–526 (1990).
122. Jungmann, R., Scheible, M. & Simmel, F. C. Nanoscale imaging in DNA nanotechnology. *Wiley Interdisciplinary Reviews: Nanomedicine and Nanobiotechnology* **4**, 66–81 (2012).
123. Henderson, R. *et al.* Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. *Journal of molecular biology* **213**, 899–929 (1990).
124. Kube, M. *et al.* Revealing the structures of megadalton-scale DNA complexes with nucleotide resolution. *Nature communications* **11**, 1–10 (2020).
125. Box, G. E. in *Robustness in statistics* 201–236 (Elsevier, 1979).
126. Kim, D.-N., Kilchherr, F., Dietz, H. & Bathe, M. Quantitative prediction of 3D solution shape and flexibility of nucleic acid nanostructures. *Nucleic acids research* **40**, 2862–2868 (2012).
127. Lee, J. Y. *et al.* Rapid computational analysis of DNA origami assemblies at near-atomic resolution. *ACS nano* **15**, 1002–1015 (2021).
128. Boltzmann, L. Studien über das Gleichgewicht der lebenden Kraft. *Wissenschaftliche Abhandlungen* **1**, 49–96 (1868).
129. Sethna, J. P. *Entropy, Order Parameters, and Complexity* (Oxford University Press, 2006).

130. Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. Equation of state calculations by fast computing machines. *The journal of chemical physics* **21**, 1087–1092 (1953).
131. Šulc, P., Doye, J. P. K. & Louis, A. A. in *Quantitative Biology, Theory, Computational Methods, and Models* (eds Munsky, B., Hlavacek, W. S. & Tsimring, L. S.) 179–208 (MIT Press, 2018).
132. Whitelam, S. & Geissler, P. L. Avoiding unphysical kinetic traps in Monte Carlo simulations of strongly attractive particles. *The Journal of chemical physics* **127**, 154101 (2007).
133. Sengar, A., Ouldridge, T. E., Henrich, O., Rovigatti, L. & Šulc, P. A Primer on the oxDNA Model of DNA: When to Use it, How to Simulate it and How to Interpret the Results. *Frontiers in Molecular Biosciences* **8** (2021).
134. Arafet, K. *et al.* Mechanism of inhibition of SARS-CoV-2 M pro by N3 peptidyl Michael acceptor explained by QM/MM simulations and design of new derivatives with tunable chemical reactivity. *Chemical science* **12**, 1433–1444 (2021).
135. Tan, D., Piana, S., Dirks, R. M. & Shaw, D. E. RNA force field with accuracy comparable to state-of-the-art protein force fields. *Proceedings of the National Academy of Sciences* **115**, E1346–E1355 (2018).
136. Sun, T. *et al.* Bottom-up coarse-grained modeling of DNA. *Frontiers in Molecular Biosciences* **8** (2021).
137. Fan, Y., Korolev, N., Lyubartsev, A. P. & Nordenskiöld, L. An advanced coarse-grained nucleosome core particle model for computer simulations of nucleosome-nucleosome interactions under varying ionic conditions. *PLoS One* **8**, e54228 (2013).
138. Savelyev, A. & Papoian, G. A. Molecular renormalization group coarse-graining of polymer chains: application to double-stranded DNA. *Biophysical journal* **96**, 4044–4052 (2009).
139. Kovaleva, N. A., Mazo, M., Zubova, E., *et al.* The “sugar” coarse-grained DNA model. *Journal of molecular modeling* **23**, 1–16 (2017).
140. Ouldridge, T. E., Louis, A. A. & Doye, J. P. Structural, mechanical, and thermodynamic properties of a coarse-grained DNA model. *The Journal of chemical physics* **134**, 02B627 (2011).
141. Šulc, P. *et al.* Sequence-dependent thermodynamics of a coarse-grained DNA model. *The Journal of chemical physics* **137**, 135101 (2012).
142. Rovigatti, L., Šulc, P., Reguly, I. Z. & Romano, F. A comparison between parallelization approaches in molecular dynamics simulations on GPUs. *Journal of computational chemistry* **36**, 1–8 (2015).
143. Snodin, B. E. *et al.* Introducing improved structural properties and salt dependence into a coarse-grained model of DNA. *The Journal of chemical physics* **142**, 06B613_1 (2015).
144. Maffeo, C. & Aksimentiev, A. MrDNA: a multi-resolution model for predicting the structure and dynamics of DNA systems. *Nucleic acids research* **48**, 5135–5146 (2020).
145. Ouldridge, T. E., Louis, A. A. & Doye, J. P. DNA nanotweezers studied with a coarse-grained model of DNA. *Physical Review Letters* **104**, 178101 (2010).
146. Srinivas, N. *et al.* On the biophysics and kinetics of toehold-mediated DNA strand displacement. *Nucleic acids research* **41**, 10641–10658 (2013).

147. Kremer, K. & Grest, G. S. Dynamics of entangled linear polymer melts: A molecular-dynamics simulation. *The Journal of Chemical Physics* **92**, 5057–5086 (1990).
148. Morse, P. M. Diatomic molecules according to the wave mechanics. II. Vibrational levels. *Physical review* **34**, 57 (1929).
149. Lennard-Jones, J. E. Cohesion. *Proceedings of the Physical Society (1926-1948)* **43**, 461 (1931).
150. Debye, P. The theory of electrolytes I. The lowering of the freezing point and related occurrences. *Physikalische Zeitschrift* **24**, 185–206 (1923).
151. Wang, Q. & Pettitt, B. M. Modeling DNA thermodynamics under torsional stress. *Biophysical journal* **106**, 1182–1193 (2014).
152. SantaLucia Jr, J. & Hicks, D. The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct.* **33**, 415–440 (2004).
153. Petukhov, M., Cregut, D., Soares, C. M. & Serrano, L. Local water bridges and protein conformational stability. *Protein Science* **8**, 1982–1989 (1999).
154. Chia-en, A. C., McLaughlin, W. A., Baron, R., Wang, W. & McCammon, J. A. Entropic contributions and the influence of the hydrophobic environment in promiscuous protein–protein association. *Proceedings of the National Academy of Sciences* **105**, 7456–7461 (2008).
155. Andersen, H. C. Molecular dynamics simulations at constant pressure and/or temperature. *The Journal of chemical physics* **72**, 2384–2393 (1980).
156. Russo, J., Tartaglia, P. & Sciortino, F. Reversible gels of patchy particles: role of the valence. *The Journal of chemical physics* **131**, 014504 (2009).
157. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *The Journal of chemical physics* **126**, 014101 (2007).
158. Doye, J. P. *et al.* The oxDNA coarse-grained model as a tool to simulate DNA origami. *arXiv preprint arXiv:2004.05052* (2020).
159. Šulc, P., Romano, F., Ouldridge, T. E., Doye, J. P. & Louis, A. A. A nucleotide-level coarse-grained model of RNA. *The Journal of chemical physics* **140**, 06B614_1 (2014).
160. Mathews, D. H. *et al.* Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proceedings of the National Academy of Sciences* **101**, 7287–7292 (2004).
161. Procyk, J., Poppleton, E. & Šulc, P. Coarse-grained nucleic acid–protein model for hybrid nanotechnology. *Soft Matter* **17**, 3586–3593 (2021).
162. LILLEY, D. J. & Clegg, R. M. The structure of the four-way junction in DNA. *Annual review of biophysics and biophysical chemistry* **22**, 299–328 (1993).
163. Mocci, F. & Laaksonen, A. Insight into nucleic acid counterion interactions from inside molecular dynamics simulations is “worth its salt”. *Soft Matter* **8**, 9268–9284 (2012).
164. Martin, T. G. & Dietz, H. Magnesium-free self-assembly of multi-layer DNA objects. *Nature communications* **3**, 1–6 (2012).

165. Snodin, B. E., Schreck, J. S., Romano, F., Louis, A. A. & Doye, J. P. Coarse-grained modelling of the structural properties of DNA origami. *Nucleic acids research* **47**, 1585–1597 (2019).
166. Snodin, B. E. *et al.* Direct simulation of the self-assembly of a small DNA origami. *ACS nano* **10**, 1724–1737 (2016).
167. Hansmann, U. H. Parallel tempering algorithm for conformational studies of biological molecules. *Chemical Physics Letters* **281**, 140–150 (1997).
168. Allen, R. J., Frenkel, D. & Ten Wolde, P. R. Forward flux sampling-type schemes for simulating rare events: Efficiency analysis. *The Journal of chemical physics* **124**, 194111 (2006).
169. Torrie, G. M. & Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics* **23**, 187–199 (1977).
170. Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H. & Kollman, P. A. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal of computational chemistry* **13**, 1011–1021 (1992).
171. Laio, A. & Parrinello, M. Escaping free-energy minima. *Proceedings of the National Academy of Sciences* **99**, 12562–12566 (2002).
172. Barducci, A., Bonomi, M. & Parrinello, M. Metadynamics. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **1**, 826–843 (2011).
173. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry* **25**, 1605–1612 (2004).
174. Suma, A. *et al.* TacoxDNA: A user-friendly web server for simulations of complex DNA structures, from single strands to origami. *Journal of computational chemistry* **40**, 2586–2595 (2019).
175. Yan, H., LaBean, T. H., Feng, L. & Reif, J. H. Directed nucleation assembly of DNA tile complexes for barcode-patterned lattices. *Proceedings of the National Academy of Sciences* **100**, 8103–8108 (2003).
176. Matthies, M. *et al.* Triangulated Wireframe Structures Assembled Using Single-Stranded DNA Tiles. *ACS Nano* **13**, 1839–1848 (2019).
177. Wei, B., Dai, M. & Yin, P. Complex shapes self-assembled from single-stranded DNA tiles. *Nature* **485**, 623–626 (2012).
178. Gopinath, A., Miyazono, E., Faraon, A. & Rothmund, P. W. Engineering and mapping nanocavity emission via precision placement of DNA origami. *Nature* **535**, 401 (2016).
179. Suma, A., Stopar, A., Nicholson, A. W., Castronovo, M. & Carnevale, V. Global and local mechanical properties control endonuclease reactivity of a DNA origami nanostructure. *Nucleic acids research* **48**, 4672–4680 (2020).
180. Maffeo, C., Yoo, J. & Aksimentiev, A. De novo reconstruction of DNA origami structures through atomistic molecular dynamics simulation. *Nucleic acids research* **44**, 3013–3019 (2016).
181. Fonseca, P. *et al.* Multi-scale coarse-graining for the study of assembly pathways in DNA-brick self-assembly. *The Journal of Chemical Physics* **148**, 134910 (2018).

182. Reinhardt, A. & Frenkel, D. Numerical evidence for nucleated self-assembly of DNA brick structures. *Physical review letters* **112**, 238103 (2014).
183. Ouldridge, T. E. *et al.* Optimizing DNA nanotechnology through coarse-grained modeling: a two-footed DNA walker. *ACS nano* **7**, 2479–2490 (2013).
184. Schreck, J. S. *et al.* DNA hairpins destabilize duplexes primarily by promoting melting rather than by inhibiting hybridization. *Nucleic Acids Research* **43**, 6181–6190 (2015).
185. Šulc, P., Ouldridge, T. E., Romano, F., Doye, J. P. & Louis, A. A. Simulating a burnt-bridges DNA motor with a coarse-grained DNA model. *Natural Computing* **13**, 535–547 (2014).
186. Doye, J. P. *et al.* Coarse-graining DNA for simulations of DNA nanotechnology. *Physical Chemistry Chemical Physics* **15**, 20395–20414 (2013).
187. Castro, C. E. *et al.* A primer to scaffolded DNA origami. *Nature methods* **8**, 221 (2011).
188. Hinckley, D. M., Freeman, G. S., Whitmer, J. K. & De Pablo, J. J. An experimentally-informed coarse-grained 3-site-per-nucleotide model of DNA: Structure, thermodynamics, and dynamics of hybridization. *Journal of Chemical Physics* **139**, 1–16 (2013).
189. Engel, M. C. *et al.* Force-induced unravelling of DNA origami. *ACS nano* **12**, 6734–6747 (2018).
190. Sharma, R., Schreck, J. S., Romano, F., Louis, A. A. & Doye, J. P. K. Characterizing the motion of jointed DNA nanostructures using a coarse-grained model. *ACS nano* **11**, 12426–12435 (2017).
191. Hong, F. *et al.* Layered-crossover tiles with precisely tunable angles for 2D and 3D DNA crystal engineering. *Journal of the American Chemical Society* **140**, 14670–14676 (2018).
192. Matek, C., Šulc, P., Randisi, F., Doye, J. P. & Louis, A. A. Coarse-grained modelling of supercoiled RNA. *The Journal of chemical physics* **143**, 243122 (2015).
193. Ouldridge, T. E., Šulc, P., Romano, F., Doye, J. P. & Louis, A. A. DNA hybridization kinetics: zippering, internal displacement and sequence dependence. *Nucleic acids research* **41**, 8886–8895 (2013).
194. Van Der Walt, S., Colbert, S. C. & Varoquaux, G. The NumPy array: A structure for efficient numerical computation. *Computing in Science and Engineering* **13**, 22–30. arXiv: 1102.1523 (2011).
195. Hunter, J. D. Matplotlib: A 2D graphics environment. *Computing in Science & Engineering* **9**, 90–95 (2007).
196. Cock, P. J. *et al.* Biopython: Freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422–1423. eprint: arXiv:1011.1669v3 (2009).
197. Pedregosa, F., Weiss, R. & Brucher, M. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **12**, 2825–2830 (2011).
198. McKerns, M. M., Strand, L., Sullivan, T., Fang, A. & Aivazis, M. A. G. Building a Framework for Predictive Science. *Proceedings of the 10th Python in Science Conference*, 1–12. arXiv: 1202.1056. <http://arxiv.org/abs/1202.1056> (2011).
199. Schickinger, M., Zacharias, M. & Dietz, H. Tethered multifluorophore motion reveals equilibrium transition kinetics of single DNA double helices. *Proceedings of the National Academy of Sciences* **115**, E7512–E7521 (2018).

200. Ester, M., Kriegel, H.-P., Sander, J., Xu, X., *et al.* A density-based algorithm for discovering clusters in large spatial databases with noise. in *Kdd* **96** (1996), 226–231.
201. Huang, C.-M., Kucinic, A., Le, J. V., Castro, C. E. & Su, H.-J. Uncertainty quantification of a DNA origami mechanism using a coarse-grained model and kinematic variance analysis. *Nanoscale* **11**, 1647–1660 (2019).
202. Baraff, D. *An introduction to physically based modeling: Rigid Body Simulation I – Unconstrained Rigid Body Dynamics* 1997.
203. Berendsen, H. J., van der Spoel, D. & van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications* **91**, 43–56 (1995).
204. Kabsch, W. A solution for the best rotation to relate two sets of vectors. *crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography* **32**, 922–923 (1976).
205. Young, G. & Householder, A. S. Discussion of a set of points in terms of their mutual distances. *Psychometrika* **3**, 19–22 (1938).
206. Borg, I. & Groenen, P. Modern multidimensional scaling: Theory and applications. *Journal of Educational Measurement* **40**, 277–280 (2003).
207. Baker, M. A. *et al.* Dimensions and Global Twist of Single-Layer DNA Origami Measured by Small-Angle X-ray Scattering. *ACS Nano* **12**, 5791–5799 (2018).
208. Schomaker, V., Waser, J., Marsh, R. E. & Bergman, G. To fit a plane or a line to a set of points by least squares. *Acta Crystallographica* **12**, 600–604 (1959).
209. David, C. C. & Jacobs, D. J. Principal Component Analysis: A Method for Determining the Essential Dynamics of Proteins. *Methods in Molecular Biology* **1084**, 193–226. <http://link.springer.com/10.1007/978-1-62703-658-0> (2014).
210. Qi, X. *et al.* Programming molecular topologies from single-stranded nucleic acids. *Nature communications* **9**, 1–9 (2018).
211. Green, A. A., Silver, P. A., Collins, J. J. & Yin, P. Toehold switches: de-novo-designed regulators of gene expression. *Cell* **159**, 925–939 (2014).
212. Pardee, K. *et al.* Rapid, low-cost detection of Zika virus using programmable biomolecular components. *Cell* **165**, 1255–1266 (2016).
213. Qi, X. *et al.* RNA origami nanostructures for potent and safe anticancer immunotherapy. *ACS nano* **14**, 4727–4740 (2020).
214. Dobrovolskaia, M. A. & Bathe, M. Opportunities and challenges for the clinical translation of structured DNA assemblies as gene therapeutic delivery and vaccine vectors. *Wiley Interdisciplinary Reviews: Nanomedicine and Nanobiotechnology* **13**, e1657 (2021).
215. Doye, J. P. *et al.* Coarse-graining DNA for simulations of DNA nanotechnology. *Physical Chemistry Chemical Physics* **15**, 20395–20414 (2013).
216. Machinek, R. R., Ouldrige, T. E., Haley, N. E., Bath, J. & Turberfield, A. J. Programmable energy landscapes for kinetic control of DNA strand displacement. *Nature communications* **5**, 5324 (2014).

217. Kočar, V. *et al.* Design principles for rapid folding of knotted DNA nanostructures. *Nature communications* **7**, 10803 (2016).
218. *ox-serve v. 1.0*; <https://doi.org/10.5281/zenodo.4551173> version 1.0. 2021. <https://doi.org/10.5281/zenodo.4551173>.
219. Kammerer, C. *et al.* Biomimetic and technomimetic single molecular machines. *Chemistry Letters* **48**, 299–308 (2019).
220. Feringa, B. L. The art of building small: from molecular switches to molecular motors. *The Journal of organic chemistry* **72**, 6635–6652 (2007).
221. Stoddart, J. F. Mechanically interlocked molecules (MIMs)—molecular shuttles, switches, and machines (Nobel Lecture). *Angewandte Chemie International Edition* **56**, 11094–11125 (2017).
222. Sauvage, J.-P. From chemical topology to molecular machines (Nobel lecture). *Angewandte Chemie International Edition* **56**, 11080–11093 (2017).
223. Bath, J. & Turberfield, A. J. DNA nanomachines. *Nanoscience And Technology: A Collection of Reviews from Nature Journals*, 124–133 (2010).
224. Erbas-Cakmak, S., Leigh, D. A., McTernan, C. T. & Nussbaumer, A. L. Artificial molecular machines. *Chemical reviews* **115**, 10081–10206 (2015).
225. Feng, Y. *et al.* Molecular pumps and motors. *Journal of the American Chemical Society* **143**, 5569–5591 (2021).
226. Von Delius, M. & Leigh, D. A. Walking molecules. *Chemical Society Reviews* **40**, 3656–3676 (2011).
227. Kamiya, Y. & Asanuma, H. Light-driven DNA nanomachine with a photoresponsive molecular engine. *Accounts of chemical research* **47**, 1663–1672 (2014).
228. Marras, A. E., Zhou, L., Su, H.-J. & Castro, C. E. Programmable motion of DNA origami mechanisms. *Proceedings of the National Academy of Sciences* **112**, 713–718 (2015).
229. Amano, S., Fielden, S. D. & Leigh, D. A. A catalysis-driven artificial molecular pump. *Nature* **594**, 529–534 (2021).
230. Feng, L. *et al.* Active mechanisorption driven by pumping cassettes. *Science* **374**, 1215–1221 (2021).
231. Erbas-Cakmak, S. *et al.* Rotary and linear molecular motors driven by pulses of a chemical fuel. *Science* **358**, 340–343 (2017).
232. Ragazzon, G., Baroncini, M., Silvi, S., Venturi, M. & Credi, A. Light-powered autonomous and directional molecular motion of a dissipative self-assembling system. *Nature nanotechnology* **10**, 70–75 (2015).
233. Kudernac, T. *et al.* Electrically driven directional motion of a four-wheeled molecule on a metal surface. *Nature* **479**, 208–211 (2011).
234. Baroncini, M. *et al.* Making and operating molecular machines: a multidisciplinary challenge. *ChemistryOpen* **7**, 169–179 (2018).
235. Wilson, M. R. *et al.* An autonomous chemically fuelled small-molecule motor. *Nature* **534**, 235–240 (2016).

236. DeLuca, M., Shi, Z., Castro, C. E. & Arya, G. Dynamic DNA nanotechnology: toward functional nanoscale devices. *Nanoscale Horizons* **5**, 182–201 (2020).
237. Zhou, L., Marras, A. E., Su, H.-J. & Castro, C. E. DNA origami compliant nanostructures with tunable mechanical properties. *ACS nano* **8**, 27–34 (2014).
238. Valero, J., Pal, N., Dhakal, S., Walter, N. G. & Famulok, M. A bio-hybrid DNA rotor–stator nanoengine that moves along predefined tracks. *Nature nanotechnology* **13**, 496–503 (2018).
239. Valero, J. & Famulok, M. Regeneration of burnt bridges on a DNA catenane walker. *Angewandte Chemie International Edition* **59**, 16366–16370 (2020).
240. Yu, Z. *et al.* A Self-Regulating DNA Rotaxane Linear Actuator Driven by Chemical Energy. *Journal of the American Chemical Society* **143**, 13292–13298 (2021).
241. Shi, Z., Castro, C. E. & Arya, G. Conformational dynamics of mechanically compliant DNA nanostructures from coarse-grained molecular dynamics simulations. *ACS nano* **11**, 4617–4630 (2017).
242. Los, G. V. *et al.* HaloTag: a novel protein labeling technology for cell imaging and protein analysis. *ACS chemical biology* **3**, 373–382 (2008).
243. Durniak, K. J., Bailey, S. & Steitz, T. A. The structure of a transcribing T7 RNA polymerase in transition from initiation to elongation. *Science* **322**, 553–557 (2008).
244. Ramezani, H. & Dietz, H. Building machines with DNA molecules. *Nature Reviews Genetics* **21**, 5–26 (2020).
245. Yoon, J., Eyster, T. W., Misra, A. C. & Lahann, J. Cardiomyocyte-driven actuation in biohybrid microcylinders. *Advanced Materials* **27**, 4509–4515 (2015).
246. Sagara, Y. *et al.* Rotaxanes as mechanochromic fluorescent force transducers in polymers. *Journal of the American Chemical Society* **140**, 1584–1587 (2018).
247. Vester, B. & Wengel, J. LNA (locked nucleic acid): high-affinity targeting of complementary RNA and DNA. *Biochemistry* **43**, 13233–13241 (2004).
248. Škugor, M. *et al.* Orthogonally photocontrolled non-autonomous DNA walker. *Angewandte Chemie* **131**, 7022–7025 (2019).
249. Wang, S. *et al.* Light-Induced Reversible Reconfiguration of DNA-Based Constitutional Dynamic Networks: Application to Switchable Catalysis. *Angewandte Chemie* **130**, 8237–8241 (2018).
250. Asanuma, H., Ito, T., Yoshida, T., Liang, X. & Komiyama, M. Photoregulation of the formation and dissociation of a DNA duplex by using the cis–trans isomerization of azobenzene. *Angewandte Chemie International Edition* **38**, 2393–2395 (1999).
251. Gerling, T. & Dietz, H. Reversible covalent stabilization of stacking contacts in DNA assemblies. *Angewandte Chemie* **131**, 2706–2710 (2019).
252. Willner, E. M. *et al.* Single-Molecule Observation of the Photoregulated Conformational Dynamics of DNA Origami Nanoscissors. *Angewandte Chemie International Edition* **56**, 15324–15328 (2017).
253. Engelhardt, F. A. *et al.* Custom-size, functional, and durable DNA origami with design-specific scaffolds. *ACS nano* **13**, 5015–5027 (2019).

254. Benson, E., Carrascosa Marzo, R., Bath, J. & Turberfield, A. J. Strategies for Constructing and Operating DNA Origami Linear Actuators. *Small* **17**, 2007704 (2021).
255. Benson, E., Marzo, R. C., Bath, J. & Turberfield, A. J. A DNA molecular printer capable of programmable positioning and patterning in two dimensions. *Science Robotics* **7**, eabn5459 (2022).
256. Rovigatti, L. *et al.* A simple solution to the problem of self-assembling cubic diamond crystals. *arXiv preprint arXiv:2205.10680* (2022).
257. Deng, Y., Tan, Y., Zhang, L., Zhang, C. & Su, X. Forecasting the Reaction of DNA Modifying Enzymes on DNA Nanostructures by Coarse Grained Model for Stimuli-Responsive Drug Delivery (2021).
258. Wang, X. *et al.* Planar 2D wireframe DNA origami. *Science advances* **8**, eabn0039 (2022).
259. Weng, Z. *et al.* Specific and robust hybridization based on double-stranded nucleic acids with single-base resolution. *Analytica Chimica Acta* **1199**, 339568 (2022).
260. Ninarello, A., Ruiz-Franco, J. & Zaccarelli, E. Onset of criticality in hyper-auxetic polymer networks. *Nature Communications* **13**, 1–8 (2022).
261. Hu, X. *et al.* Structure-Guided Designing Pre-Organization in Bivalent Aptamers. *Journal of the American Chemical Society* **144**, 4507–4514 (2022).
262. Stephanopoulos, N. Hybrid nanostructures from the self-assembly of proteins and DNA. *Chem* **6**, 364–405 (2020).
263. Dass, M., Gür, F. N., Kołataj, K., Urban, M. J. & Liedl, T. DNA origami-enabled plasmonic sensing. *The Journal of Physical Chemistry C* **125**, 5969–5981 (2021).
264. Dong, Y. & Mao, Y. DNA Origami as Scaffolds for Self-Assembly of Lipids and Proteins. *ChemBioChem* **20**, 2422–2431 (2019).
265. Kut'ák, D., Poppleton, E., Miao, H., Šulc, P. & Barišić, I. Unified nanotechnology format: one way to store them all. *Molecules* **27**, 63 (2021).
266. Monticelli, L. *et al.* The MARTINI coarse-grained force field: extension to proteins. *Journal of chemical theory and computation* **4**, 819–834 (2008).
267. Naskar, S. & Maiti, P. K. Mechanical properties of DNA and DNA nanostructures: comparison of atomistic, Martini and oxDNA models. *Journal of Materials Chemistry B* **9**, 5102–5113 (2021).
268. Thomen, P. *et al.* T7 RNA polymerase studied by force measurements varying cofactor concentration. *Biophysical journal* **95**, 2423–2433 (2008).
269. Poppleton, E., Romero, R., Mallya, A., Rovigatti, L. & Šulc, P. OxDNA. org: a public webserver for coarse-grained simulations of DNA and RNA nanostructures. *Nucleic acids research* **49**, W491–W498 (2021).
270. Centola, M. *et al.* A rhythmically pulsing leaf-spring nanoengine that drives a passive follower. *bioRxiv* (2021).

APPENDIX A

SUPPLEMENTARY MATERIAL FOR CHAPTER 2

A.1 OxDNA Input and Output File Formats

As the analyses discussed in the main text process the data from oxDNA files, it warrants a brief description of the file types encountered when running and analyzing oxDNA files. This will mitigate confusion as to why certain files are used in each case.

The three most important files are the input parameter file, the topology file and the trajectory file. The input file is a text file with the simulation's parameter names and values separated by equal signs. This file defines simulation and physical parameters such as temperature, the force field used, and file I/O information. Input files will be used by any analysis that calculates energy (primarily those concerned with hydrogen bonds) or is optimized by writing its most computationally-intensive operations distributed and compiled with the oxDNA code.

Topology files define the nucleotides present in the simulation. It is a text file with the extension .top and contains a header line with the number of nucleotides and the number of strands separated by a space. Following the header, each line defines a nucleotide with four values: the strand number, the base identity, 3' covalent connection, and 5' covalent connection (note that oxDNA numbers bases 3'-5', the reverse of the biochemistry convention). Strand ends have a value of -1 .

The trajectory file contains the position, orientation and velocity of every nucleotide at each timepoint, based on the interval specified in the input file. Each configuration in the trajectory begins with a three-line header, defining the temperature, simulation box dimensions, and kinetic, potential, and total energies of the system. Each subsequent line defines one nucleotide with 15 parameters: position, orientation defined by two orthogonal vectors, translational velocity and rotational velocity. All parameters are in XYZ coordinates. The TacoxDNA webserver¹⁷⁴ has a variety of conversion tools from popular nanotechnology design tools and simulation formats into the oxDNA format. Once converted, these files can be visualized and edited using oxView or simulated using oxDNA/oxRNA. However, for large and/or long simulations, these files can become quite unwieldy, requiring tens of gigabytes of storage and therefore are impossible to open and read in a single reading frame. Therefore, all analyses and visualizations described in the main text read trajectory files in a stream, allowing reading of files that would not otherwise fit in the computer's RAM.

A.2 Simulation Details

The simulations given as examples here were run using the oxDNA code (June 2019 version). Structures were originally obtained in either Tiamat⁹⁴ or CaDNAno⁹⁷ format and then exported to oxDNA format using the TacoxDNA webserver¹⁷⁴. Structures were relaxed in two steps. First, a brief Monte-Carlo (MC) simulation was performed using the DNA_relax or RNA_relax force fields to remedy overlapping particles. After this initial relaxation, mutual traps based on the intended design were applied to enforce relaxation to the intended design. For designs exported from Tiamat, mutual trap files are produced by TacoxDNA; for other structures, these files were produced using the force file generation script described in the "other utilities" section after the MC relax. A further relaxation was performed using the max_backbone_force option in a molecular dynamics (MD) simulation with the DNA2 or RNA2 force field. This bypasses checks of backbone bond length and allows for faster relaxation due to the CUDA implementation of the MD method in oxDNA¹⁴². This simulation was run until the energy stabilized between -57.98 and -62.13 pN nm (-1.4 and -1.5 oxDNA energy units respectively). At which point, the external forces and backbone force limitations were released and a production simulation run was performed using the same force field for 10^9 steps with a stepsize of 15.15 fs (0.005 oxDNA time units). This corresponds to a total run time in the microsecond range; however, previous work with the oxDNA model^{166,189} suggests, in part, due to the increased diffusion coefficient, this direct conversion is an underestimate of the corresponding experimental time. However, as is the problem with all coarse-grained models, it is impossible to establish a direct correspondence between the simulation and experimental time because different processes in a coarse-grained simulation can scale to the experiment with different ratios.

The production simulations were performed at 20° C using an Andersen-like thermostat¹⁵⁶, and configurations were saved for analysis every $5 * 10^5$ steps, resulting in 2000 separate configurations used in each analysis.

The RNA tile, used as an example of the clustering algorithm, was initially run as described above, however simulation runs would frequently transition to a state where one of the crossovers was broken. Further simulation only sampled the broken state and failed to capture reversible transition. Assessing the free-energy difference between states through simulation requires observing the transition multiple times. To facilitate observing the transition, a parallel tempering simulation using virtual-move Monte Carlo (VMMC)¹³² with umbrella sampling was performed. Simulations with eight parallel replicas were performed with the temperatures set from 25 to 60° C at 5 degree increments. Replica exchange was attempted every 1000 steps and the average exchange acceptance rate was 0.47. Achieving reversible transitions along this order parameter also required creation of a weight file to bias the simulation towards transition states between the two states. This simulation successfully sampled multiple transitions between the two states and the resulting combined trajectory was used to demonstrate the clustering script.

A.3 RNA Tile Analysis with Unsupervised Clustering

The trajectory files from all replicates of an RNA tile simulation discussed in the main text were combined into a single trajectory file containing 1345 configurations. The output from the principal component script was fed into the DBSCAN clustering algorithm as implemented in python sci-kit with parameters $\text{eps} = 12$ and $\text{min_samples} = 8$. This generated 5 different clusters of configurations, of which the first three are displayed in Fig. 11 of the main text. The number of configurations in each cluster were 812, 240, 110, 78 and 7 with 98 unclustered configurations. It is possible to have fewer configurations in a cluster than min_samples because min_samples only sets the number of neighbors required for a single point to be in a cluster. The last two clusters correspond to partially melted states stabilized by oxRNA's extremely strong cross-stacking interactions. We do not believe that these states are physically relevant, however they are successfully separated out by the clustering algorithm. The positions of each configuration projected to the first three axes in principal component space are shown in Fig. A1.

A.4 Interduplex Angles and Distances

Figures A2 and A3 illustrate examples of analysis of interduplex angles and distances between origami units respectively, as discussed in the main text.

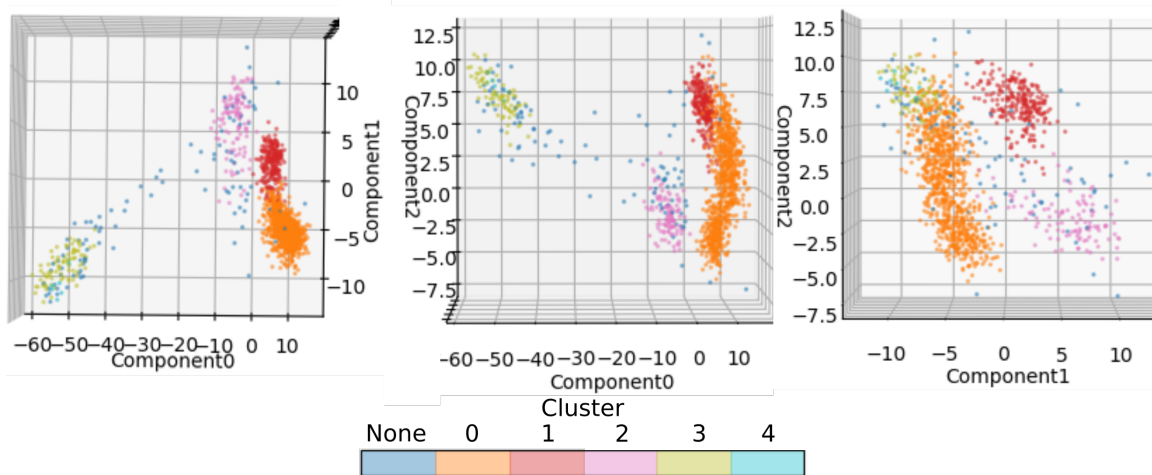


Figure A1. The positions of each configuration projected to the first three dimensions of principal component space. The three plots correspond to looking down components 2, 1, and 0, respectively. Clusters 0, 1 and 2 correspond to the designed tile structure, the designed structure with stacking interrupted at the nick point, and the structure where the paranemic cohesion is lost becoming a Holliday junction. clusters 3 and 4 correspond to non-physically relevant states where the entire strand becomes a single hairpin.

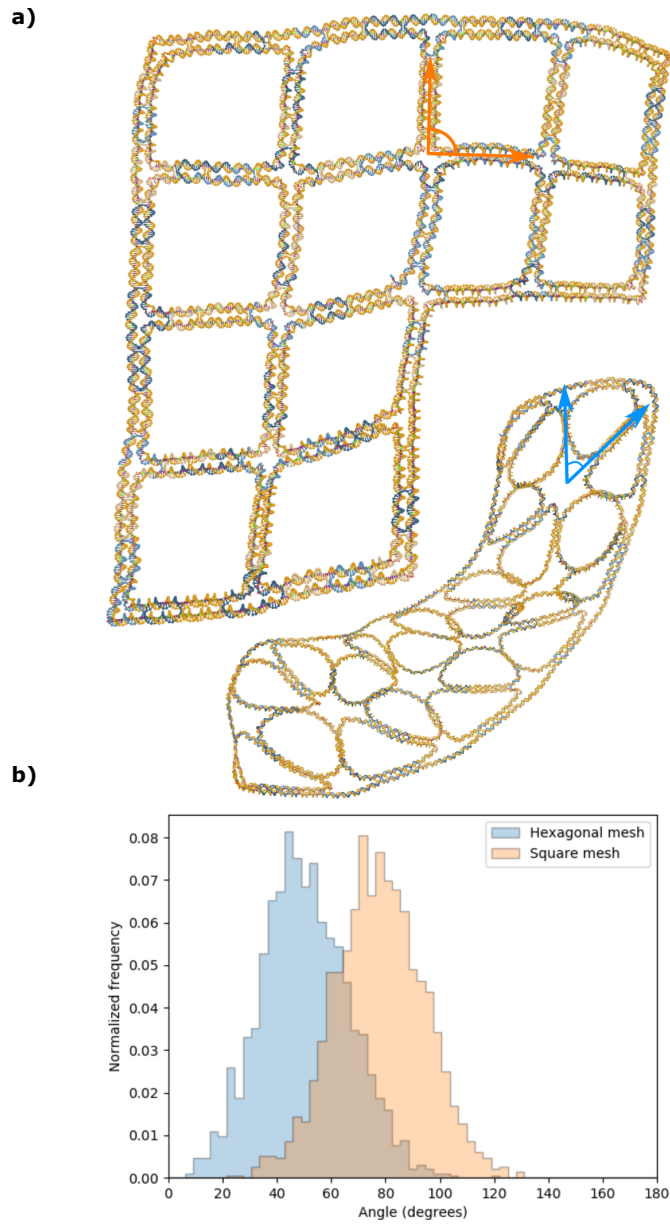


Figure A2. **Comparing angles in wireframe lattices.** **a)** The mean structures of design 23 (top) and design 20 (bottom) from¹⁰⁸. The structures are designed to have a square and hexagonal lattice pattern, respectively. **b)** The distribution of angles between two arms of a junction showing variation around the designed junction angle. For the hexagonal lattice, the observed angle is lower than the designed angle of 60° because the structure has significant out-of-plane curvatures in the simulation.

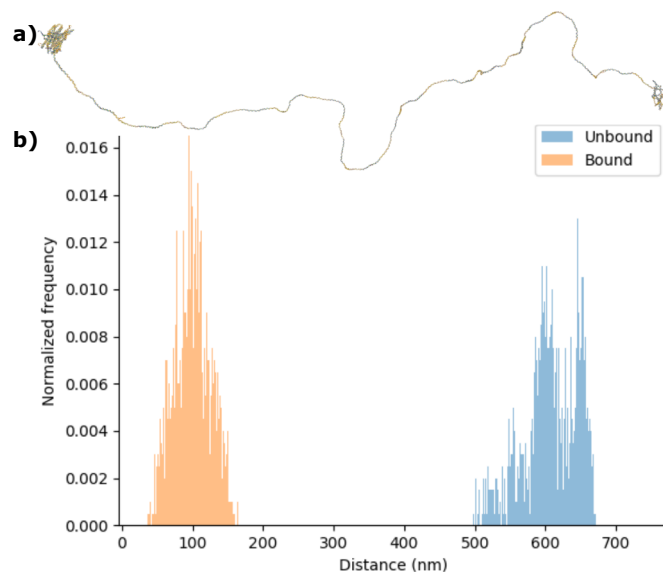


Figure A3. **Distance between origami units of TMF.** **a)** The final configuration from a simulation of the TMF structure used in DNA kinetics experiments¹⁹⁹. Separate simulations were performed with the sticky ends in both the bound and free configurations. **b)** The distribution of distances between the origami units at opposite ends of the tether.



Figure A4. **3D printed Holliday junction exported from oxView.** OxView supports export to GLTF format that can be opened in a 3D rendering tool Blender and exported to 3D printers or used for creation of more artistic 3D figures of DNA and RNA nanostructures.

APPENDIX B

SUPPLEMENTARY MATERIAL FOR CHAPTER 4

B.1 Materials and Methods

B.1.1 Design and Assembly of the 18-Helix Bundle DNA Origami

The origami was designed around the M13mp18 circular single stranded DNA scaffold of 7249 nucleotides in length utilizing the three-dimensional DNA designing tool Cadnano2 (<https://cadnano.org>). Staples have been optimized to have an average length and when possible optimized to follow design strategies presented in previous publications that described the MgCl_2 free assemblies of DNA origamis.¹⁶⁴ Three-dimensional representation of the design was obtained with the online Tool CanDo (<https://cando-dna-origami.org/>). The tool can provide 3D structure predictions based on the Cadnano2 designs. The provided rmsf 3D maps have been used to optimize the structure to maximise the overall origami stability. From the optimized designs, atomic models are obtained from the CanDo tool and further used to confirm shape and dimension of the designed structure and are used to find the more suitable places where further modification can be inserted. The place to introduce the chloroalkane modified staple, that serves as attachment point for the Halo T7 RNAP fusion protein, and the transcribable dsDNA strand are introduced in the more stable origami parts. The stiff origami arms are the most stable at half length of the origami arm. In order to not reduce the overall stability of the origami, we have chosen to introduce modifications to the overall structure into staples that already are nicked and are in the correct position where the modifications are needed. To ensure that the modification protrude orthogonally to the surface of the origami, the neighbouring staple is lengthened and made complementary to the protruding sequence of the modified oligo. The short double-stranded sequence that protrudes from the origami ensures that the sequence is normal to the origami structure as long as the sequence is much shorter than the persistence length of DNA. Origamis are assembled by combing 13.3 nM of the M13mp18 scaffold with 10 EQ of each staple in 1 x OB. The mixture is then divided into 50 μl aliquots into 500 μl reaction tubes and the origami annealed in a thermocycler.

B.1.2 Molecular Dynamics Simulations

Multiple designs of the NE were simulated using the oxDNA coarse grain model for DNA origami. OxDNA is described in detail elsewhere¹⁵⁸, but briefly, it is an empirically-derived force field designed with DNA nanostructures in mind. The model has been shown to reproduce structural, kinetic and thermodynamic properties of DNA, including persistence length, strand displacement rates and free energy barriers between states, with reasonably high accuracy, while still being sufficiently coarse-grained to allow simulations of DNA origami at timescales of up to milliseconds¹⁶⁶. Equilibrium simulations of the NE were carried out for 3 different designs: nNE, NE, and a version of nNE with two staples removed from the hinge (soft nicked NE, snNE). Starting configurations were obtained by exporting the CaDNAno⁹⁷ design file into oxDNA format using an in-house conversion script. Rigid body dynamics in oxView visualization tool^{114,118} were used to bend the arms into a starting configuration. Relaxation was then performed using the method described in¹⁵⁸. After relaxation, the bridge was built using oxView's editing tools and a further round of relaxation performed. After the average energy per particle stabilized around -1.5 su, the structures were equilibrated with production conditions for a further $2.5e8$ simulation steps to allow equilibration of the angle distribution (step number determined by past simulations). Equilibrium simulations were carried out for $1e9$ oxDNA simulation steps with a timestep of integration of 0.003, a temperature of either 23° C or 37° C was imposed using an Andersen-like thermostat. Configuration snapshots were saved for analysis every $5e5$ steps giving 2000 configurations per simulation which were verified to be well decorrelated. To identify the effects of secondary structure in the flexure region, a second set of equilibrium simulations was performed with the base type of the nucleotides in the single-stranded region of the flexure set to non-interacting (no structure, NS). These simulations were carried out with the same parameters as the simulations where secondary structure was allowed to form. Closing rates were measured by running either 3 (snNE) or 10 (nNE, nNE-NS, NE and NE-NS) replicates of simulations with external forces added. The simulations were started from the final

configuration of one replicate of the associated equilibrium simulation and run for $1e7$ steps with snapshots saved every $1e3$ steps for a total of 2000 configurations per simulation. A constant force of 16 pN (based on the typical value of tension exerted by a polymerase on a duplex DNA²⁶⁸) was applied between the nucleotide where the RNA polymerase is covalently linked to the first nucleotide in the first stop sequence in the bridge with a cutoff radius of 10.35 nm (the radius of T4 polymerase)²⁴³. Other parameters were the same as in the equilibrium simulations. One additional simulation was performed for each design where the force was applied for $1e9$ steps to allow the structure to equilibrate in the closed position. Opening rates were measured by running either 3 (snNE) or 10 (nNE, nNE-NS, NE and NE-NS) replicates starting from the final configuration of the equilibrated pulling simulation. Each simulation was run for $2e7$ steps with snapshots saved every $1e3$ steps for a total of 4000 configurations per simulation. In addition to the full NE, simulations were also performed where the bridge was deleted (NE-NB and NE-NS-NB), allowing the structure to open under only the influence of the flexure. Other parameters were the same as in the equilibrium simulations. Simulations were aligned and mean structures obtained from the equilibrium simulations using `oxdna_analysis_tools`¹¹⁴ and movies of the trajectories produced using `oxView`^{114,118}. To get the spring constant and opening/closing rates, a linear regression was performed on the point clouds corresponding to the arms of each hinge and the angle between the arms calculated for each snapshot. The spring constant was then estimated using the equipartition theorem for a simple harmonic oscillator

$$k(x - x_0)^2 = \frac{1}{2}K_bT \quad (\text{B.1})$$

Where k is the spring constant, x and x_0 are the displacement and the average position, and K_bT is the thermal energy. Opening and closing rates were calculated using a linear regression to the angle traces over the simulations. Significance of distributions was determined using a two-tailed Kolmogorov-Smirnov test.

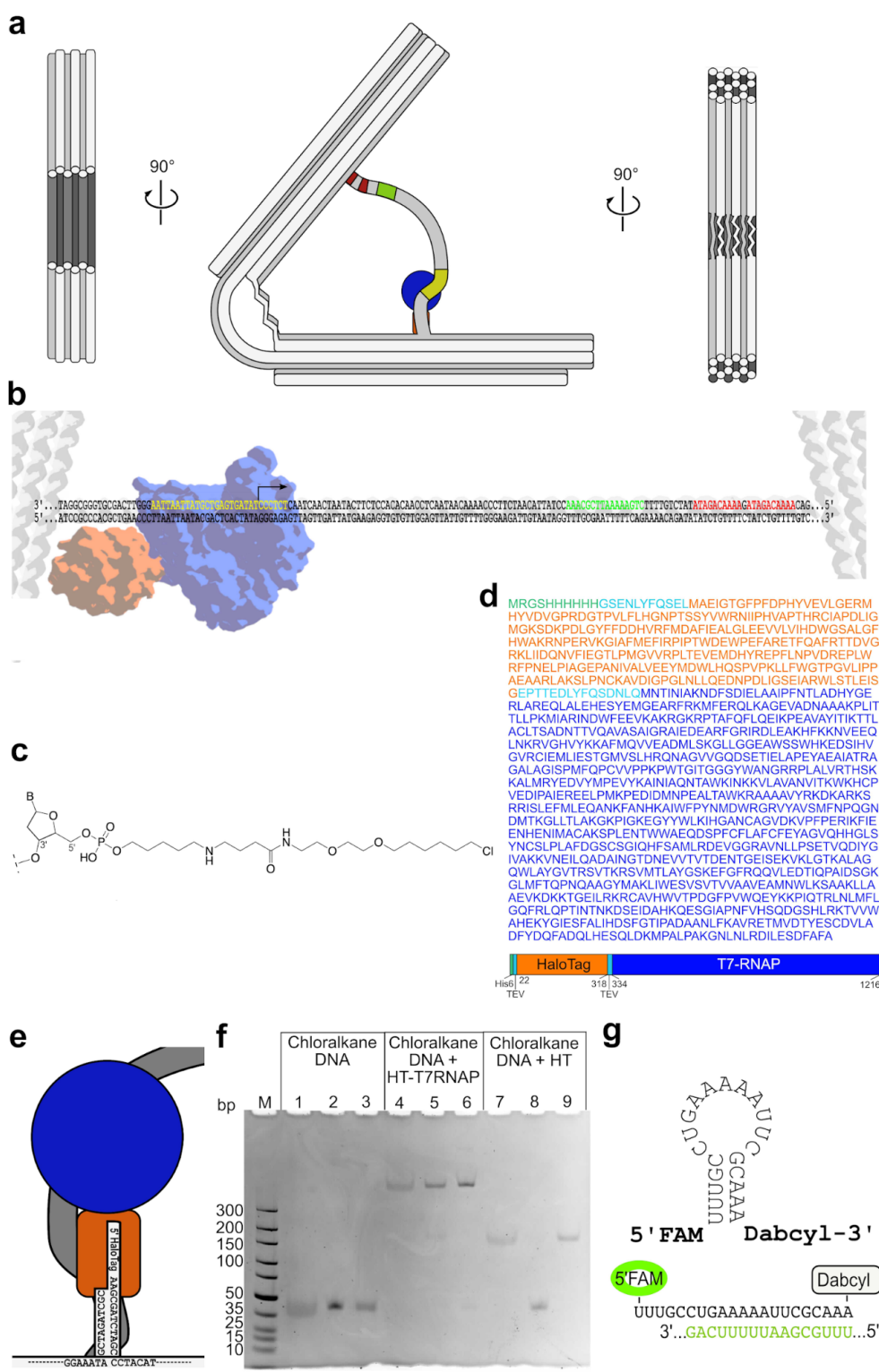


Figure B1. (Previous page) **Additional design features of the leaf-spring NE** **(a)** Cross section of the NE with the leaf-spring (left panel), a side-view (middle) that shows the ssDNA template and the location of the T7 promotor (yellow), the sequence coding for a binding site for a molecular beacon to determine transcription yields (green), and the two terminator sequences (red). The right panel shows the location of the six ssDNA sequences, **(b)** Sequence of the dsDNA transcription template. Yellow: T7 promotor, green: sequence coding for molecular beacon-binding (green), red: terminator sequence, **(c)** Chemical structure of the halogenated 5'-end of the protruding HT-T7RNAP attachment staple, **(d)** Primary amino acid sequence (upper panel) and design (lower panel) of the HT-T7-RNAP fusion protein, **(e)** Design and sequence of the protruding staple containing the 5'-halogenated attachment site for HT-T7RNAP, **(f)** Chloroalkane DNA connection to the HaloTag (HT) enzyme. Lanes 1-3: Chloralkane DNA in (1) origami buffer, (2) H₂O, (3) Origami buffer + EDTA; lanes 4-6: Chloralkane DNA + HT-T7RNAP in (4) origami buffer, (5) H₂O, (6) Origami buffer + EDTA; lanes 7-9: Chloralkane DNA + HT in (7) origami buffer, (8) H₂O, (9) Origami buffer + EDTA. The Halo enzyme alone in absence of MgCl₂ cannot bind to the chloroalkane modified DNA. In presence of the buffer the protein can bind to the DNA and the connection is covalent and strong enough that even after addition of EDTA to remove the MgCl₂ the connection of DNA and protein is maintained. In the fusion protein of Halo T7 RNA pol this phenomenon is not present, very likely due to the affinity of the polymerase towards DNA that probably increases the affinity of the fused Halo tag towards the DNA, **(g)** Sequence, secondary structure, and labels (5'-FAM, 3'-Dabcyl) of the molecular beacon RNA that detects the green sequence in the RNA generated during transcription.

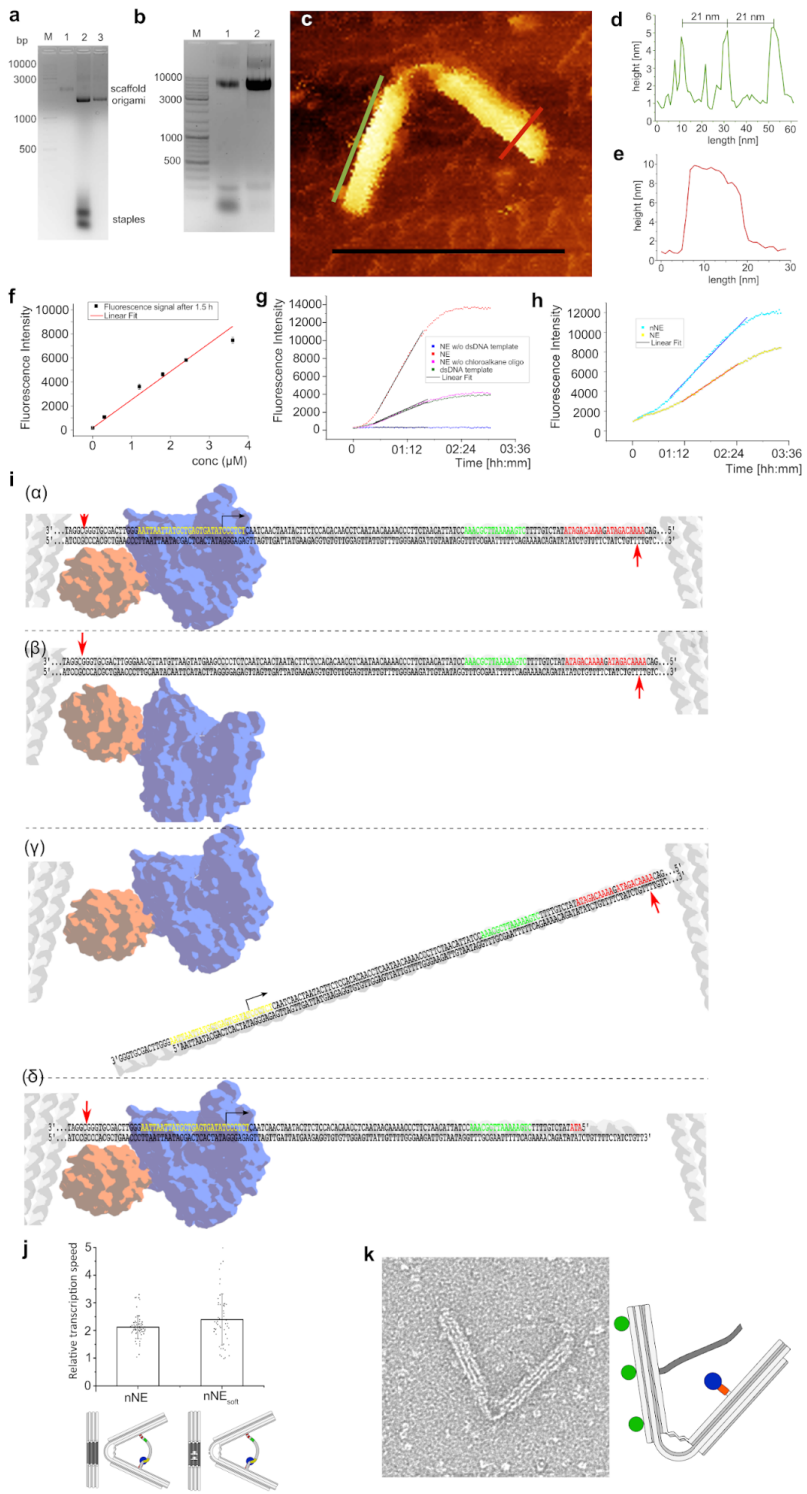


Figure B2. (Previous page) **Assembly of the NE and AFM analysis of the geometry of the NE.** (a) Filter purification of the origami. Lane 1: M13MP18 scaffold, lane 2: assembled origami unpurified, lane 3: 100 kDa filter-purified origami, (b) PEG-purification of origami. Lane 1: Assembled unpurified origami, lane 2: PEG-purified origami, (c) detailed AFM image of the origami structure. The green line spans the streptavidin molecules attached to the respective origami-arm and marks the height measurement shown in (d). The red line marks the cross-section of the opposing origami-arm and marks the height measurement shown in (e), (d) height profile of the green line shown in (c) confirms the distance of the streptavidin molecules on the respective origami-arm to be spaced exactly 21 nm as designed, (e) height profile of the red line shown in (c) to determine the cross-section of the opposing origami-arm to be exactly 9 nm as designed, (f) calibration curve of the fluorescence intensity (F.I.) of the molecular beacon (MB) as a function of the concentration of the added complementary oligonucleotide. The measuring time was 1.5 h because the F.I. stabilized after that time without significant further photobleaching, (g) exemplary transcription curves of constructs 1a (blue), 2a (green), 3a (magenta), and 4a (red) shown in Fig. 2a with linear fit of the linear parts of the curves, (h) exemplary transcription curves of constructs 1b (yellow) and 2b (cyan) with linear fit of the linear parts of the curves, (i) Detailed representation of the various forms of template dsDNA used in this study; (α) red arrows: position of the two nicks in the template dsDNA, (β) red arrows: the same without the promoter region, (γ) dsDNA, attached only on the opposite side of the HT-T7RNAP; red arrow: nick-position, (δ) dsDNA, attached only next to the HT-T7RNAP; red arrow: nick position. Importantly, to avoid that the single stranded nicks weakens the stability of the template DNA and cause detachment of the DNA from the origami by forces generated during the pulling by the polymerase and/or during the formation of the transcription bubble, the template strand has no single nicks upstream of the promoter region while the nick is placed in the coding strand. Downstream of the promoter region the single stranded nick has been placed into the template strand while the coding strand is fully attached into the origami. This design avoids that the two ss nicks in the coding strand lead to loss of the DNA-DNA duplex by formation of a DNA-RNA as the amount of RNA increases over time. Both, leading and coding strands, are anchored to the origami; their hybridization is favored even when the RNA transcript levels increase, (j) Relative transcription speed of nNE (left) and nNEsoft (right). Error bars: S.D., $n \geq 53$. (k) representative TEM image (upper panel) of construct 1d indicates electrostatic repulsion of the dsDNA template strand in case where the strand is not anchored at its anchoring point next to the HT-T7-RNAP, which hampers efficient transcription. Lower panel: cross section of this construct.

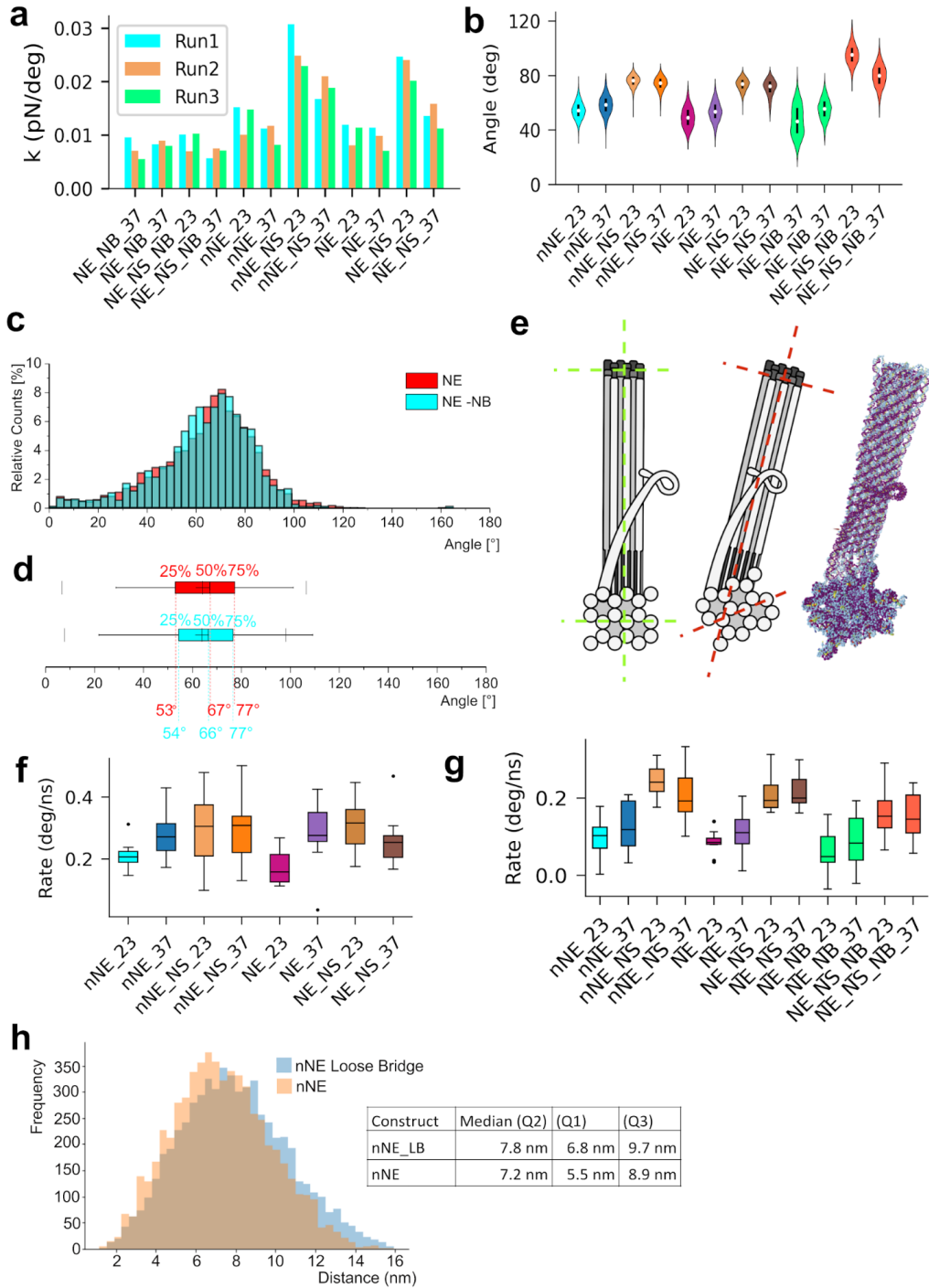


Figure B3. (Previous page) **Comparison of hinge properties of simulations run at 23° and 37° C.** **(a)** Calculated spring constant of each design at 23° and 37° C. Only NS structures with an intact dsDNA bridge show a significant difference between the two temperatures, demonstrating the influence of stable secondary structures on angle distribution, **(b)** Angle-distribution in triplicate equilibrium simulations at different temperatures. A slight increase in average angle for simulations with secondary structures in the flexure and slight decrease in average angle for simulations where secondary structure was inhibited was observed. This is consistent with the hypothesis that secondary structures limit the opening angle of the structures with secondary structure while the extended single strands behave more like entropic springs which become more flexible as temperature increases. **(c)** The angle distributions between NE and NE-NB are highly comparable ($p = 0.6$). The box plot in **(d)** which shows distribution of angles measured from TEM images of NE (red, $n = 5135$) with the angle distribution of the NE-NB missing the dsDNA template strand (cyan, $n = 1382$) with whiskers of size 1.5 times with first, second (median) and third quartile indicated with dashed lines (red for NE and cyan for NE-NB). Thin cross: average values $64^\circ \pm 20^\circ$ for NE and $64^\circ \pm 19^\circ$ for NE-NB. **(e)** Relaxation of the NE origami structure simulation indicate that the origami arms bent out of plane. Left panel: designed structure, middle panel: model based on simulation, right panel: relaxed simulation. The origami arms clearly show a longitudinal twist in the top view of the 18 HB and are bent out of the vertical plane of the arm that extends towards the top. Since the dsDNA helices in the origami arms are used as references, a distortion of the surface-deposited structure results in slight systematic difference compared to how the angles are measured in the simulated structure. This difference likely explains the systematic difference in the angle distributions. **(f)** Calculated pulling rates under 16 pN applied force between the polymerase attachment point and the terminator sequence on the dsDNA bridge for each design at 23° and 37° C. The only significant difference between the two temperatures was observed for nNE and NE, where decreasing the number of base pairs in the flexure significantly increased the rate at which the leaf-spring was able to close, **(g)** Calculated re-opening rates after being closed under 16 pN force. This is a process which relies mostly on brownian motion to return to the relaxed state of the NE. Unsurprisingly, we see an increase in average rate and an increase in variance for the structures where secondary structures can form in the flexure and a decrease in average rate and no trend in variance for structures where no structure was permitted in the flexure. **(h)** Distance distribution between the polymerase attachment point and start of the promoter sequence. The radius representing the attached polymerase has been estimated to be 10.35 nm to have an encounter of promoter region and polymerase. The frequency at which the nNE construct (red curve, nNE) spends within this ideal distance bubble is higher than the case where the transcribable dsDNA strand is attached only next to the polymerase (blue curve, nNE_LB).

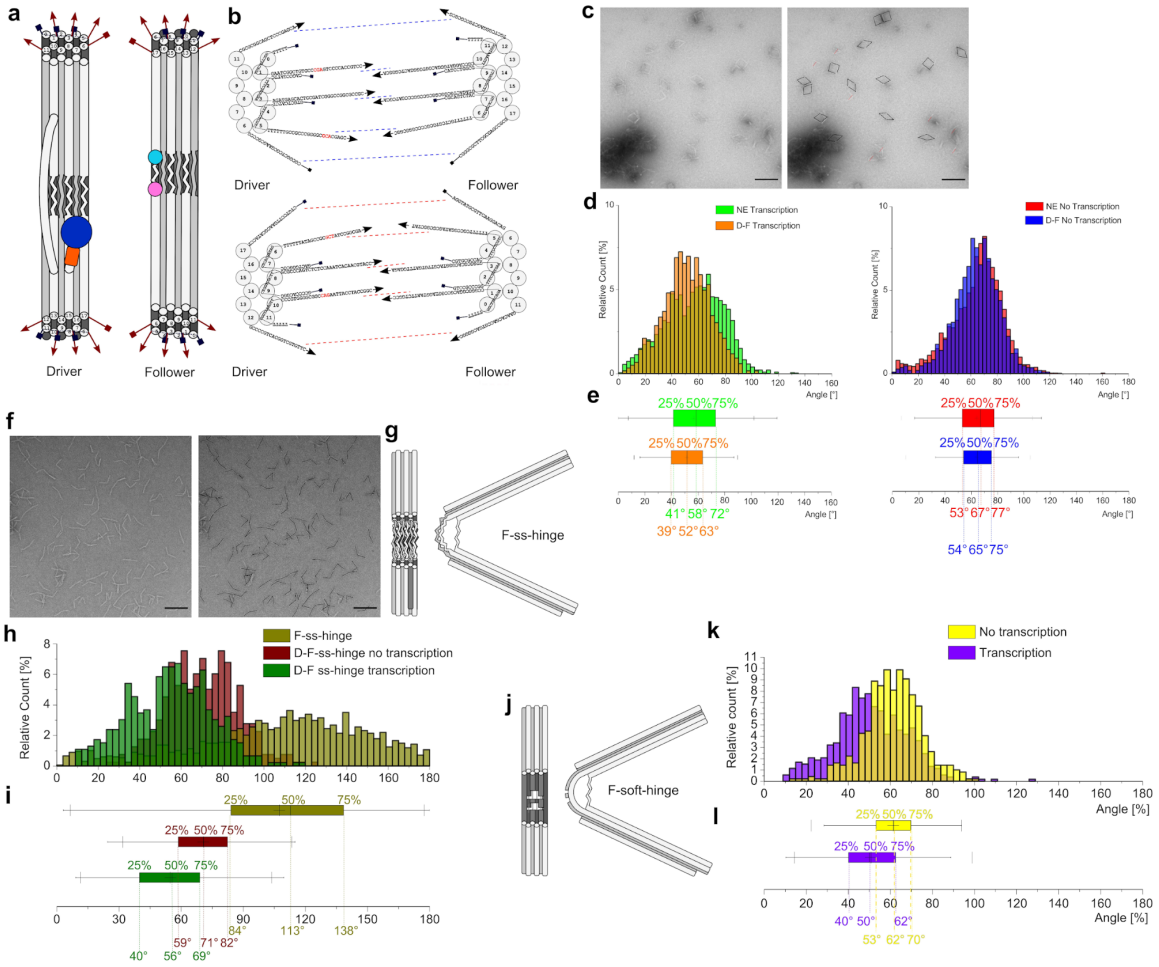


Figure B4. (Previous page) **Driver follower experiments** Schematic front view of the Driver and Follower units, **(a)** Note how it is necessary to flip one of the two units upside down to be able to join the structures. **(b)** DNA sequences of the overhangs for the Driver (left) and for the Follower (right). Bases labeled in red in the Driver overhang sequences indicate LNA. **(c)** shows an example of TEM image used to determine the angle of the D-F complex on the left and with the overlaid black line to indicate how the angle were measured. Improperly joined structures or single origamis have been ignored during the angle measurements (red strikethrough). When measuring the angle distribution from TEM images **(d)** and comparing the distribution of NE (red, $n = 5135$) with the angle distributions of only complete D-F (blue, $n = 1074$) in absence of transcription it is noticeable how the average angle remains unchanged, being $64^\circ \pm 20^\circ$ for NE and $64^\circ \pm 17^\circ$ (Error: S.D.) but the curve is less skewed and more symmetric in case of the D-F complex. The same effect is appreciable also during transcription (NE in green, $n = 3266$; D-F in orange, $n = 1190$). In the NE sample the average angle distribution drops to $57^\circ \pm 22^\circ$ that is comparable to the average value of the D-F complex of $57^\circ \pm 17^\circ$ but the curve is narrower in the second case. The boxplots of the distribution, **(e)**, confirm that in case of the D-F complex the angle distribution is narrower and less skewed. **(e)**. The example TEM image for the F-no-hinge structure **(f)** nicely shows how large the range of angles is when the double stranded structure is missing in the flexure region making it impossible to obtain the correct structure that determines the angulated form of the origami structure. The F-no-hinge **(g)** constructs shows a very wide and flat distribution **(h)** (dark yellow, $n = 1682$) with obtuse angle of $107^\circ \pm 41^\circ$. When combining the F-no-hinge origami with the D unit the angle is reduced to $71^\circ \pm 17^\circ$ with a narrower distribution (wine, $n = 398$). In case of transcription the average distribution angle shifts to $55^\circ \pm 20^\circ$ getting slightly larger as expected for the transcription sample (olive, $n = 462$). The Boxplots **(i)** show clearly how the distribution for the F-no-hinge (dark yellow) is widely spread over a great range of angles, having a difference between Q3 and Q1 of 54° while in the D-F this range is reduced to 23° in absence of transcription (wine) and 29° in presence of transcription (olive). The change in angle distribution during transcription in the D-F structure is also appreciable in the presence of F-soft-hinge **(j)** in the complex **(k)**. The D-F-soft-hinge complex shows an average angle of $61^\circ \pm 14^\circ$ (yellow, $n = 496$) in the no transcription case and $51^\circ \pm 17^\circ$ in the transcription sample (violet, $n = 528$). The boxplot **(l)** graphically shows how also in this case the distribution clearly shifts towards more acute angles in the case of transcription compared to the no transcription sample. (Errors are S.D. and the boxplots show Q1, Q2, Q3 in the box with whiskers 1.5 times the box size. Thin crosses in the boxplots indicate the average angle while the thin vertical lines indicate the 1% and 99% percentile of the distribution).

APPENDIX C

PLEASE SEE ATTACHED FILES

C.1 Supplementary Videos


1. Video 1: Lemniscate video of oxView scene containing more than 1 million nucleotides
2. Video 2: OxView editing demonstration
3. Video 3: Rigid body dynamics in oxView
4. Video 4: Lemniscate video of a mean structure
5. Video 5: Video of simulation trajectory

APPENDIX D

PERMISSIONS FOR MANUSCRIPT RE-USE

The co-authors of the papers reproduced for chapters 2¹¹⁴, 3²⁶⁹ and 4²⁷⁰ have given their permission for the manuscripts to reappear here.

The published manuscripts for chapters 2¹¹⁴ and 3²⁶⁹ are under a creative commons license and do not require permissions to reproduce as long as they are properly attributed. Chapter 4 is a preprint²⁷⁰, so the copyright is held by the authors.



Design, optimization and analysis of large DNA and RNA nanostructures through interactive visualization, editing and molecular simulation

Author: Poppleton, Erik; Bohlin, Joakim
Publication: Nucleic Acids Research
Publisher: Oxford University Press
Date: 2020-05-25


Copyright © 2020, Oxford University Press

Creative Commons

This is an open access article distributed under the terms of the [Creative Commons CC BY](#) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

You are not required to obtain permission to reuse this article.

Figure D1. License Information for Chapter 2



OxDNA.org: a public webserver for coarse-grained simulations of DNA and RNA nanostructures

Author: Poppleton, Erik; Romero, Roger
Publication: Nucleic Acids Research
Publisher: Oxford University Press
Date: 2021-05-01

Copyright © 2021, Oxford University Press

Creative Commons

This is an open access article distributed under the terms of the [Creative Commons CC BY](#) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.


You are not required to obtain permission to reuse this article.

Figure D2. License Information for Chapter 3

APPENDIX E

PERMISSIONS FOR ADAPTED FIGURES

a



Enhanced Performance of 50 nm Ultra-Narrow-Body Silicon Carbide MOSFETs based on FinFET effect
 Conference Proceedings: 2020 32nd International Symposium on Power Semiconductor Devices and ICs (ISPSD)
 Author: T. Kato
 Publisher: IEEE
 Date: Sept. 2020
 Copyright © 2020, IEEE

Thesis / Dissertation Reuse

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:

- 1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
- 2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
- 3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:

- 1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
- 2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
- 3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

b

Publisher: Elsevier
 Copyright © 1969, Elsevier

Creative Commons

This is an open access article distributed under the terms of the [Creative Commons CC-BY](#) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

You are not required to obtain permission to reuse this article.

To request permission for a type of use not listed, please contact [Elsevier Global Rights Department](#).

Are you the author of this Elsevier journal article?

Figure E1. Licensing information for parts (a) and (b) of Figure 1.

License Number	5327830900668		
License date	Jun 14, 2022		
Licensed Content		Order Details	
Licensed Content Publisher	Elsevier	Type of Use	reuse in a thesis/dissertation
Licensed Content Publication	Materials Science and Engineering: C	Portion	figures/tables/illustrations
Licensed Content Title	Sonochemical in situ immobilization of Pd nanoparticles on green tea extract coated Fe ₃ O ₄ nanoparticles: An efficient and magnetically recyclable nanocatalyst for synthesis of biphenyl compounds under ultrasound irradiations	Number of figures/tables/illustrations	1
Licensed Content Author	Hojat Veisi, Milad Ghorbani, Saba Hemmati	Format	both print and electronic
Licensed Content Date	May 1, 2019	Are you the author of this Elsevier article?	No
Licensed Content Volume	98	Will you be translating?	No
Licensed Content Issue	n/a		
Licensed Content Pages	10		
About Your Work		Additional Data	
Title	Software tools for design, simulation and characterization of DNA and RNA nanostructures	Portions	Figure 2
Institution name	Arizona State University		
Expected presentation date	Jun 2022		
Requestor Location		Tax Details	
	Erik Poppleton 10436 N Central Ave	Publisher Tax ID	98-0397604
Requestor Location	PHOENIX, AZ 85020 United States Attn: Erik Poppleton		
Price			
Total	0.00 USD		

Figure E2. Re-use permission for part (c) of Figure 1

License Number	5327840099499		
License date	Jun 14, 2022		
Licensed Content		Order Details	
Licensed Content Publisher	Elsevier	Type of Use	reuse in a thesis/dissertation
Licensed Content Publication	Food Research International	Portion	figures/tables/illustrations
Licensed Content Title	Characterization and stability evaluation of β -carotene nanoemulsions prepared by high pressure homogenization under various emulsifying conditions	Number of figures/tables/illustrations	1
Licensed Content Author	Yuan Yuan, Yanxiang Gao, Jian Zhao, Like Mao	Format	both print and electronic
Licensed Content Date	Jan 1, 2008	Are you the author of this Elsevier article?	No
Licensed Content Volume	41	Will you be translating?	No
Licensed Content Issue	1		
Licensed Content Pages	8		
About Your Work		Additional Data	
Title	Software tools for design, simulation and characterization of DNA and RNA nanostructures	Portions	Figure 3a
Institution name	Arizona State University		
Expected presentation date	Jun 2022		
Requestor Location		Tax Details	
	Erik Poppleton 10436 N Central Ave	Publisher Tax ID	98-0397604
Requestor Location	PHOENIX, AZ 85020 United States Attn: Erik Poppleton		
Price			
Total	0.00 USD		

Figure E3. Re-use permission for part (d) of Figure 1

License Number	5327981323719		
License date	Jun 14, 2022		
Licensed Content		Order Details	
Licensed Content Publisher	Springer Nature	Type of Use	Thesis/Dissertation
Licensed Content Publication	Nature Biotechnology	Requestor type	academic/university or research institute
Licensed Content Title	GPSeq reveals the radial organization of chromatin in the cell nucleus	Format	print and electronic
Licensed Content Author	Gabriele Girelli et al	Portion	figures/tables/illustrations
Licensed Content Date	May 25, 2020	Number of figures/tables/illustrations	1
		High-res required	no
		Will you be translating?	no
		Circulation/distribution	1 - 29
		Author of this Springer Nature content	no
About Your Work		Additional Data	
Title	Software tools for design, simulation and characterization of DNA and RNA nanostructures	Portions	4a
Institution name	Arizona State University		
Expected presentation date	Jun 2022		
Requestor Location		Tax Details	
	Erik Poppleton 10436 N Central Ave		
Requestor Location	PHOENIX, AZ 85020 United States Attn: Erik Poppleton		
Price			
Total	0.00 USD		

Figure E4. Licensing information for parts (a) of Figure 2

License Number	5327990168188		
License date	Jun 14, 2022		
Licensed Content		Order Details	
Licensed Content Publisher	Elsevier	Type of Use	reuse in a thesis/dissertation
Licensed Content Publication	Cell	Portion	figures/tables/illustrations
Licensed Content Title	Ribosome Structure and the Mechanism of Translation	Number of figures/tables/illustrations	1
Licensed Content Author	V. Ramakrishnan	Format	both print and electronic
Licensed Content Date	Feb 22, 2002	Are you the author of this Elsevier article?	No
Licensed Content Volume	108	Will you be translating?	No
Licensed Content Issue	4		
Licensed Content Pages	16		
About Your Work		Additional Data	
Title	Software tools for design, simulation and characterization of DNA and RNA nanostructures	Portions	1b
Institution name	Arizona State University		
Expected presentation date	Jun 2022		
Requestor Location		Tax Details	
	Erik Poppleton 10436 N Central Ave	Publisher Tax ID	98-0397604
Requestor Location	PHOENIX, AZ 85020 United States Attn: Erik Poppleton		

Figure E5. Re-use permission for part (b) of Figure 2

License Number	5327990736644		
License date	Jun 14, 2022		
Licensed Content		Order Details	
Licensed Content Publisher	Elsevier	Type of Use	reuse in a thesis/dissertation
Licensed Content Publication	Trends in Biochemical Sciences	Portion	figures/tables/illustrations
Licensed Content Title	How Myosin Generates Force on Actin Filaments	Number of figures/tables/illustrations	1
Licensed Content Author	Anne Houdusse,H. Lee Sweeney	Format	both print and electronic
Licensed Content Date	Dec 1, 2016	Are you the author of this Elsevier article?	No
Licensed Content Volume	41	Will you be translating?	No
Licensed Content Issue	12		
Licensed Content Pages	9		
About Your Work		Additional Data	
Title	Software tools for design, simulation and characterization of DNA and RNA nanostructures	Portions	2
Institution name	Arizona State University		
Expected presentation date	Jun 2022		
Requestor Location		Tax Details	
	Erik Poppleton 10436 N Central Ave	Publisher Tax ID	98-0397604
Requestor Location	PHOENIX, AZ 85020 United States Attn: Erik Poppleton		
Price			
Total	0.00 USD		

Figure E6. Re-use permission for part (c) of Figure 2

License Number	5330020747556		
License date	Jun 15, 2022		
Licensed Content		Order Details	
Licensed Content Publisher	Springer Nature	Type of Use	Thesis/Dissertation
Licensed Content Publication	Nature	Requestor type	academic/university or research institute
Licensed Content Title	Diverse and robust molecular algorithms using reprogrammable DNA self-assembly	Format	print and electronic
Licensed Content Author	Damien Woods et al	Portion	figures/tables/illustrations
Licensed Content Date	Mar 20, 2019	Number of figures/tables/illustrations	1
		High-res required	no
		Will you be translating?	no
		Circulation/distribution	1 - 29
		Author of this Springer Nature content	no
About Your Work		Additional Data	
Title	Software tools for design, simulation and characterization of DNA and RNA nanostructures	Portions	3b
Institution name	Arizona State University		
Expected presentation date	Jun 2022		
Requestor Location		Tax Details	
	Erik Poppleton 10436 N Central Ave		
Requestor Location	PHOENIX, AZ 85020 United States Attn: Erik Poppleton		
Price			
Total	0.00 USD		

Figure E7. Re-use permission for part (b) of Figure 3

License Number	5330021073951		
License date	Jun 15, 2022		
Licensed Content		Order Details	
Licensed Content Publisher	Springer Nature	Type of Use	Thesis/Dissertation
Licensed Content Publication	Nature	Requestor type	academic/university or research institute
Licensed Content Title	Folding DNA to create nanoscale shapes and patterns	Format	print and electronic
Licensed Content Author	Paul W. K. Rothemund	Portion	figures/tables/illustrations
Licensed Content Date	Mar 16, 2006	Number of figures/tables/illustrations	1
		High-res required	no
		Will you be translating?	no
		Circulation/distribution	1 - 29
		Author of this Springer Nature content	no
About Your Work		Additional Data	
Title	Software tools for design, simulation and characterization of DNA and RNA nanostructures	Portions	2
Institution name	Arizona State University		
Expected presentation date	Jun 2022		
Requestor Location		Tax Details	
	Erik Poppleton 10436 N Central Ave		
Requestor Location	PHOENIX, AZ 85020 United States Attn: Erik Poppleton		
Price			
Total	0.00 USD		

Figure E8. Re-use permission for part (c) of Figure 3

License Number	5330040009586		
License date	Jun 15, 2022		
Licensed Content		Order Details	
Licensed Content Publisher	Springer Nature	Type of Use	Thesis/Dissertation
Licensed Content Publication	Nature Materials	Requestor type	academic/university or research institute
Licensed Content Title	Programmable icosahedral shell system for virus trapping	Format	print and electronic
Licensed Content Author	Christian Sigl et al	Portion	figures/tables/illustrations
Licensed Content Date	Jun 14, 2021	Number of figures/tables/illustrations	1
		High-res required	no
		Will you be translating?	no
		Circulation/distribution	1 - 29
		Author of this Springer Nature content	no
About Your Work		Additional Data	
Title	Software tools for design, simulation and characterization of DNA and RNA nanostructures	Portions	5e
Institution name	Arizona State University		
Expected presentation date	Jun 2022		
Requestor Location		Tax Details	
	Erik Poppleton 10436 N Central Ave		
Requestor Location	PHOENIX, AZ 85020 United States Attn: Erik Poppleton		
Price			
Total	0.00 USD		

Figure E9. Re-use permission for part (d) of Figure 3