DETC2023-113633

AN IMPROVED SHAPE ANNEALING ALGORITHM FOR THE GENERATION OF COATED DNA ORIGAMI NANOSTRUCTURES

Bolutito Babatunde
Department of
Mechanical Engineering
Carnegie Mellon
University
Pittsburgh, PA

Jonathan Cagan
Department of
Mechanical Engineering
Carnegie Mellon
University
Pittsburgh, PA

Rebecca E. Taylor
Department of
Mechanical Engineering
Carnegie Mellon
University
Pittsburgh, PA

ABSTRACT

In recent years, the field of structural DNA nanotechnology has advanced rapidly due to transformative design tools. Although these tools have been revolutionary, they still bear one overall limitation of requiring users to fully conceptualize their designs before designing. Recently, a simple computational casting technique was developed using generative optimization strategies to automate the design of DNA origami nanostructures. This approach employs a shape annealing algorithm, which creates a formal language of honeycomb DNA origami nanostructures with shape grammars and drives designs from the language towards a desired configuration using simulated annealing. This initial demonstration of the approach can generate novel scaffold routing schemes for creating solid or hollow structures constrained by boundaries of polyhedral meshes. The results from the initial approach, particularly from the hollow structures, reveal a challenging design space. This simple technique generates novel scaffold routing schemes that do not replicate the overall polyhedral mesh shape and is limited in its ability of controlling scaffold path exploration in the design space. This paper demonstrates an approach for varying effective wall thickness and improving quality of polyhedral mesh coverage for hollow structures that can be tuned and optimized by introducing a more refined computational casting technique. We achieve these improvements through changes in the simulated annealing algorithm by adding a Hustin move set algorithm that dynamically adjusts the performance of the overall design and redefining how these hollow designs are articulated. The results in this work illustrate how the shape annealing algorithm can navigate a challenging design space to generate effective hollow designs.

Keywords: design optimization, generative design, micro and nano systems design and synthesis of, biomaterials, design automation, structural DNA nanotechnology, and DNA origami

NOMENCLATURE

|	
bp	base pair
f(x)	objective function
M	mutations or iterations
L	limit
T	temperature
T_p	previous temperature
T_i	initial temperature
α	alpha (reduction factor)
r	shape rule
Q_r	quality factory for the rth shape rule
n_r	number of times the rth shape rule is called at T_p
s	total number of shape rules
$P_{i,r}$	initial probability if the rth shape rule
P_u	probability multiplier
$prob_i$	initial probability for the Hustin move set
Q_{total}	total quality factor over all shape rules
x	minimum perpendicular Euclidean distance from the outer wall of input mesh to the center of the 7 bp scaffold section
d	set desired distance from the outer wall of input mesh

1. INTRODUCTION

In the nanoworld, there has been an exponential increase in the development of tools for the design of structures and machines using DNA nanotechnology [1–10]. These approaches are based on Watson-Crick base pairing in DNA, wherein adenine (A) and guanine (G) pair with thymine (T) and cytosine (C) via hydrogen bonds [11–13]. Figure 1 shows a 3D molecular representation of a 7-base pair (bp) single helix section of a DNA double helix with simplifications where a single helix can be represented using ball-and-stick models or cylindrical building blocks. By harnessing the programmable binding of nucleobases, the users can control the assembly of biomechanical nanostructures [14]. The most common method for forming complex DNA nanostructures is an approach called DNA origami. As shown in Figure 2a, this self-assembly method utilizes a long single-stranded DNA (ssDNA), also known as a "scaffold", that is folded and then cinched together with hundreds of oligonucleotides, also known as "staples", to form a desired configuration [12,15]. A common scaffold used in the field of DNA origami is M13mp18 (M13), which is obtained from a bacteriophage and is 7249 bases in length [16]. The DNA origami designs can be represented in 2D schematics and 3D renderings in which the scaffold and staples are simplified to cylinders (Figure 2b). The DNA origami folding method can create multilayer or wireframe architectures with a myriad of applications including nanosensors [17], nanolithography [18], nanomachines [19], and nanocasting [20].
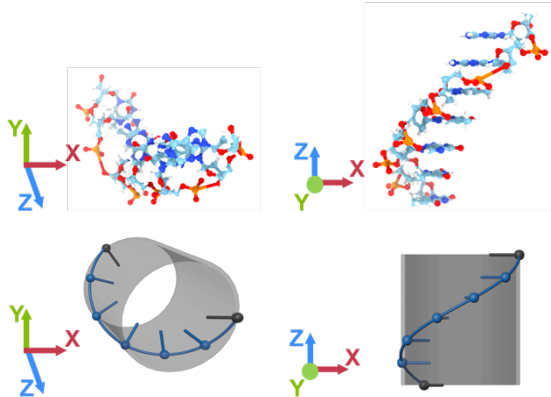
FIGURE 1: REPRESENTATION OF 7 BP SCAFFOLD SEGMENT: (TOP) HELIX REPRESENTATION. (BOTTOM) SIMPLIFIED REPRESENTATION.

Computational casting is a new approach that uses the shape annealing generative algorithm [21] to automate the routing of the scaffold strand around or within an input polyhedral mesh, resulting in coated-type or solid designs, respectively [22]. The shape annealing algorithm uses shape grammars [23] to create a language of DNA nanostructures or shape rules that defines the path through space, and simulated annealing [24] drives the language towards an optimally preferred configuration. This technique paves the way for novel design tools in which humans are not required to define the scaffold routing for irregular multilayer DNA origami nanostructures. By using a complex polyhedral mesh as a design boundary, this approach adds a

multilayer functionality that complements existing automated tools for generating 2D and 3D wireframe DNA origami nanostructures [3,5–10]. Here, polyhedral mesh refers to a simple geometric object. This method is particularly groundbreaking in the nanoworld because the generated scaffold routes are capable of being optimized for structure and function. It is important to note that the polyhedral mesh is merely a boundary that is removed after the scaffold is generated.
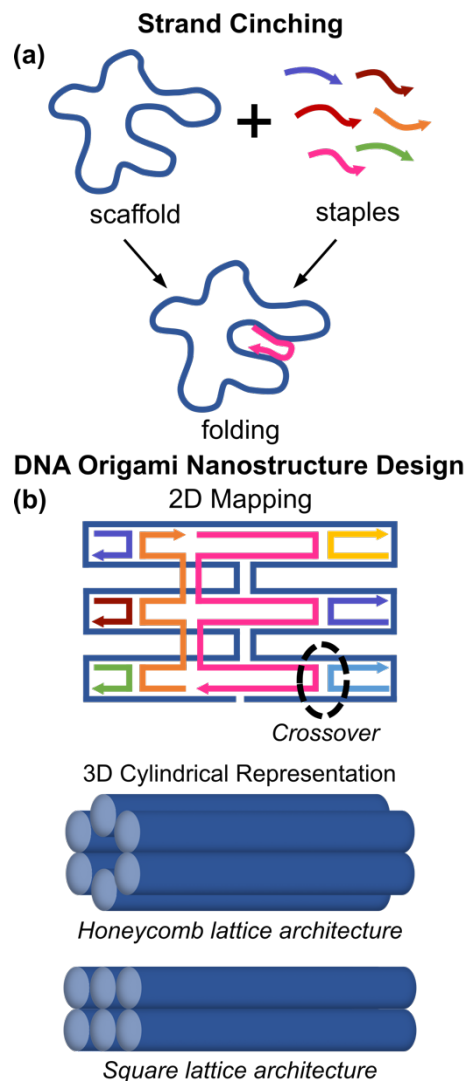
FIGURE 2: DNA ORIGAMI SELF-ASSEMBLY METHOD: (A) SCAFFOLD FOLDING (B) DNA ORIGAMI NANOSTRUCTURE DESIGN VISUALIZATION.

In this work we address the quality and thickness of generated designs using a refined version of our computational casting technique to route the scaffold around a tetrahedral mesh (Figure 3). This has not been demonstrated in prior work. The refined technique demonstrates that the objective function from the simulated annealing algorithm can tune the mean effective thickness of the coated-type designs. Therefore, this approach can achieve desired coatings with a variety of scaffold lengths.

This feature is of the utmost importance to the DNA nanotechnology community because the scaffold material length is often a key constraint for design, limiting the size of what can be built.
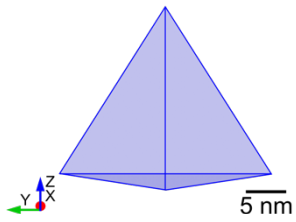
FIGURE 3: INPUT TETRAHEDRAL MESH FOR REFINED AND SIMPLE TECHNIQUE WITH AXIS ALIGNED BOUNDING BOX (L, W, H) DIMENSIONS OF (21.7, 25.0 ,20.4) NM.

This work also investigates the density within the walls for various thicknesses, determining that internal porosity or quality can also be tuned by the objective function. Such an approach which has not been addressed with prior algorithms shows how a design algorithm applied to this new problem can have a profound impact in the output quality. This paper addresses performance limitations in previous work for coated-type designs which was apparent in large gaps within the walls and low wall density where the overall structure did not replicate the shape of the tetrahedral mesh (Figure 4) [22]. Figure 4 displays an example of the cross-sectional view generated with the simple algorithm from previous work. In the figure, the rectangle to the left highlights the location of the cross-sectional view on the overall structure. This same cross-sectional representation is maintained for the refined technique. Such limitations show that the design problem of coating a scaffold around irregular geometries creates a challenging design space that cannot be tackled with simple optimization strategies. To tackle the challenging design space and further improve the quality of the coated designs, this work appends the simulated annealing algorithm with the Hustin move set [25]. This technique dynamically adjusts the probability of the shape rules to improve the performance of the algorithm, tune the wall thickness, and decrease the large gaps observed in previously generated designs.
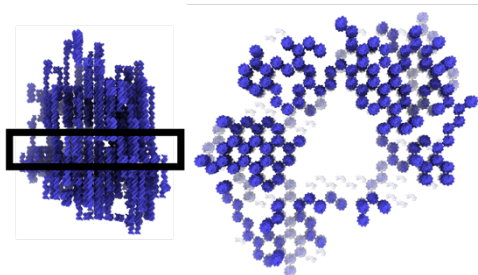
FIGURE 4: EXAMPLE OF GENERATED DNA ORIGAMI DESIGN USING SIMPLE TECHNIQUE WITH TOTAL SCAFFOLD LENGTH OF 12,901 BP.

2. MATERIALS AND METHODS

This section provides a brief description of the generative methods in the computational casting approach from previous work and details the refinements made to tune the mean effective thickness of generated coated-type DNA origami designs. The same three shape grammar rules are used, and a refined simulated annealing algorithm is then used to control the formation of these structures to meet the well-defined coating criteria. In the simulated annealing algorithm, coating behavior is rearticulated by replacing the objective function from previous work. The Hustin move set is then introduced as a modification to the simulated annealing algorithm to iteratively increase the probability of selected shape rules that help boost the algorithm's coating performance.

2.1 Shape Grammars

In shape annealing, a shape grammar is a set of 2D or 3D shape rules that are applied sequentially from a starting geometric shape to generate a set of designs or language [23], while simulated annealing controls the evolution of the starting shape. In this case, the geometric shape takes on key features of a single-stranded DNA (ssDNA) in 3D form with a helical diameter of 2 nm and an axial rise per base of 0.34 nm or roughly 34.3º turn per base. The shape grammar rules in this work mirror the rules from the previous computational casting technique [22], which adopt the honeycomb lattice architecture from the popular manual computer-aided design (CAD) tool, caDNAno [2]. Since the honeycomb lattice permits a union (or crossover) between neighboring helices at multiples of 7 bases at roughly 240º, the design language divides the scaffold into 7 bp sections ensuring continuous scaffold growth through the addition of a new section.
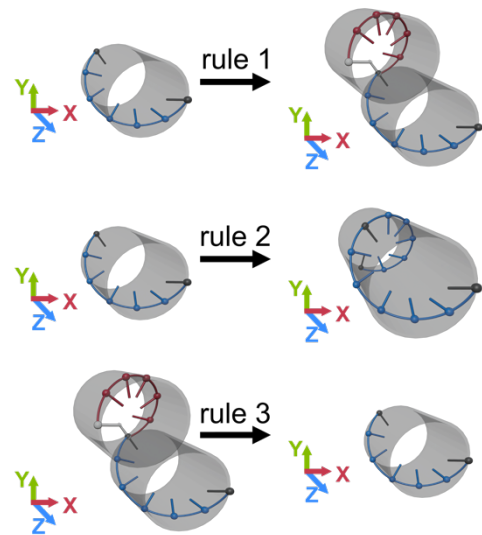
FIGURE 5: SHAPE GRAMMAR

Figure 5 illustrates simple examples of the three-rule shape grammar used in the previous work. The blue single helix shows helix growth in the +z-direction while the red single helix shows helix growth in the -z direction. The first rule is the crossover rule which generates a 7 bp increment of scaffold on the nearest of the three possible neighboring helices at angles 0º, 120º, and 240º. Figure 5 shows the crossover at 120º. The second rule is

the extension rule which extends the scaffold by a 7 bp section along the z-axis of its current helix. Figure 5 shows extension of the blue single helix. The third rule is the reversal rule which undoes a previous call of the first and second rule to stop the scaffold from being stuck and encourages scaffold exploration within the given design space. Figure 5 shows the undoing of Rule 1. The rules create a language or set of instructions that drive the path of a continuous scaffold strand along a caDNAno honeycomb lattice architecture.

2.2 Optimization with Simulated Annealing

Shape annealing is a version of simulated annealing which is a stochastic optimization technique, where each state is calculated using a randomly generated shape rule as a move set. In previous work, the simulated annealing algorithm has an outer and inner loop. Within the inner loop of M mutations or iterations, each randomly generated feasible state is assessed using an objective function, $f(x)$. For objective minimization, if $\Delta f(x) \leq 0$, where the evaluated current and previous states are compared, the current state is accepted. However, if the condition is false, a probability of adding the state is calculated:

$$P_{add} = e^{-\frac{\Delta f(x)}{T}}, \qquad (1)$$

where $\Delta f(x)$ is the change in objective function and T is the temperature. The current state can only be added if a normalized random number is less than the calculated probability. If a set limit (L) of added states has been reached, the algorithm jumps out of the inner loop. In the outer loop, if no states have been added then the algorithm has reached equilibrium and can be terminated. If the condition is false, the temperature is lowered with an exponential function:

$$T = \alpha T_p, \qquad (2)$$

where T_p is the previous temperature value and α, which is between 0 and 1, is the reduction factor. The algorithm runs until either equilibrium is reached or the temperature reaches zero.

2.2.1 Hustin Move Set

In this paper, we apply an additional step within the outer loop of the simulated annealing algorithm called the Hustin move set, which adjusts the probability of selecting a given rule based on the performance of the algorithm at any time [25]. The rationale for dynamically modifying the probabilities using the Hustin move set is that the refined technique will increase the probability of selecting the appropriate shape rule for an increased coverage and wall density. This therefore decreases the number of poor shape rule acceptances made during the anneal. This common technique has been used to address the problem of optimal nonorthogonal routing of components in a chemical plant [26]. This problem is similar to the DNA scaffold routing problem in this work where the Hustin move set is used to modify the probability of selecting each shape rule. Previously, the computational casting technique set an equal probability of selecting each shape rule throughout an anneal. The generated results from this fixed shape rule selection probability display

sparse scaffold coverage of input polyhedral mesh and low wall density in the overall structure (Figure 4). Based on analysis, we hypothesize these performance limitations to originate from a lower call in the extension shape rule (rule 2) due to a higher ratio of neighboring helical sections of only 7 bp in length (rule 1) within regions of sparse scaffold coverage. Therefore, modification of shape probability selection is necessary.

In the modified simulated annealing algorithm, the probabilities of implementing moves are updated after each temperature reduction as a function of the success of the moves (or shape rules) in accomplishing the objective at the previous temperature. The measure of success per shape rule (r) is determined by a quality factor,

$$Q_r = \frac{1}{n_r} \sum_{\substack{rules \\ accepted}} |\Delta f(x)|, \qquad r = 1, \dots s, \qquad (5)$$

where n_r is the number of times the rth shape rule was called at the previous temperature, $\Delta f(x)$ is the change in objective due to an accepted shape rule from objective minimization, and s is the total number of shape rules. If a shape rule is not called at a specific temperature value, then its quality factor is set to zero. At the start of an anneal, each shape rule probability is initialized to ensure that each rule is likely to be called. While this initial probability could be skewed with an initial bias or uniform for all shape rules, the sum of all initial probabilities must equal 1:

$$\sum_{\substack{intial \\ probabilities}} P_{i,r} = 1, \qquad (6)$$

where $P_{i,r}$ is the initial probability of the rth shape rule. Here, the initial probability for the extension rule is much higher than the crossover and reversal rules. Once the quality factor is calculated, the probability of selecting the rth shape rule is updated with the following equation:

$$P_r = P_{i,r} P_u + (1 - P_u) \frac{Q_r}{Q_{total}}, \qquad r = 1, \dots s, \qquad (7)$$

where P_u, which is between 0 and 1, is a multiplier that guarantees the relevance of the initial bias per temperature reduction and Q_{total} is the total quality factor over all shape rules. A shape rule with a larger quality factor will have a higher call probability at the next temperature.

2.3 Shape Annealing

Shape annealing uses simulated annealing to control when a randomly selected shape rule (or scaffold section) is accepted to the overall continuous scaffold routing pattern at a given state. Of note, prior uses of shape annealing in design focused on the design of truss structures [27]. One contribution of this work is the extension and demonstration of the algorithm applied at the nano scale to design DNA structures.

Figure 6 illustrates a flow chart of the revised shape annealing casting algorithm including the extension to include the Hustin move set. In the flowchart, there are three key inputs: starting shape, specification, and objective and constraints. The starting shape, which is a match from the previous technique, is a blue scaffold 7 bp scaffold section. The specifications are the same key shape annealing parameters from previous work with an appended modification: initial temperature (T_i), limit (L),

mutations (M), reduction factor (α), and initial probability ($prob_i$) for Hustin move set. The constraints and objectives mirror the previous technique, where the constraint given is the tetrahedral mesh as a boundary and the objective is for scaffold to wall distance minimization. The algorithm focuses on shape rule transformation where each shape rule is evaluated based on the refined simulated annealing algorithm. In the flow chart the crossover rule (rule 1) is selected and applied to the starting shape which consists of a single blue segment which is transformed to a blue and red 14 bp long scaffold. The shape rule probabilities are updated per temperature reduction and the image for this step is an example where the cylinders in red are the accepted shape rules used to calculate the new probabilities. Once the overall shape has reached convergence, the shape annealing algorithm will stop and the output will be the final design.

2.4 Structural Validation with Coarse-Grained Simulations

Validating the shape and stability of these generated complex DNA origami designs is essential for ensuring structure formation during the time-intensive physical experimentation. There are a variety of established structural validation tools, which model the traits of DNA origami through computer simulations that provide detailed predictions of the behavior of the DNA nanostructures comparable to the results from physical experimentation [15,28–35]. OxDNA is a coarse-grained DNA model that is used as a tool to capture the structural, mechanical, and thermodynamic characteristics of DNA at the nucleotide resolution level over timescales ranging from, but not limited to, nanoseconds to microseconds that are relevant for studies of nanostructure behavior [32–34]. OxDNA provides invaluable information about the local and global dynamic properties of these DNA nanostructures such as the response of the structure due to internal stress [35], yielding under tension [36], and actuation [37].
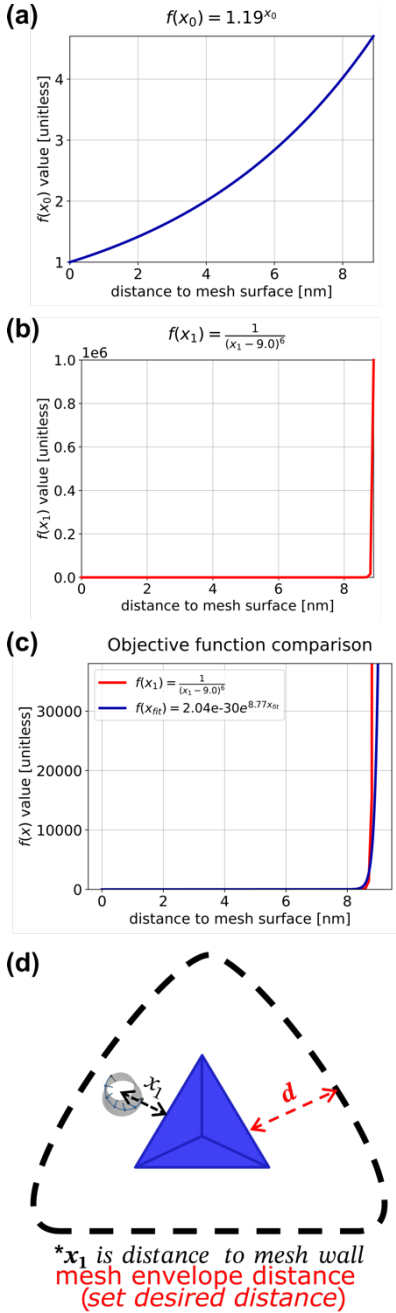
This work uses oxDNA to model the generated DNA origami designs following a common relaxation process [31] in which overlapping nucleotide volumes and stretched backbone bonds are corrected. The general goal of the relaxation process is to reach a steady state through energetic minimization of the system. The relaxation process is two-part, where a minimization algorithm runs for 2000 steps and a relaxation molecular dynamics simulation run for 10^7 steps. After the DNA origami nanostructure is well relaxed, another molecular dynamics simulation step is subsequently run for 10^7 steps with the absence of added forces or modified potentials used for root mean-square fluctuations (RMSF) calculations. After running simulations, these structures are analyzed using a standardized set of tools [38].
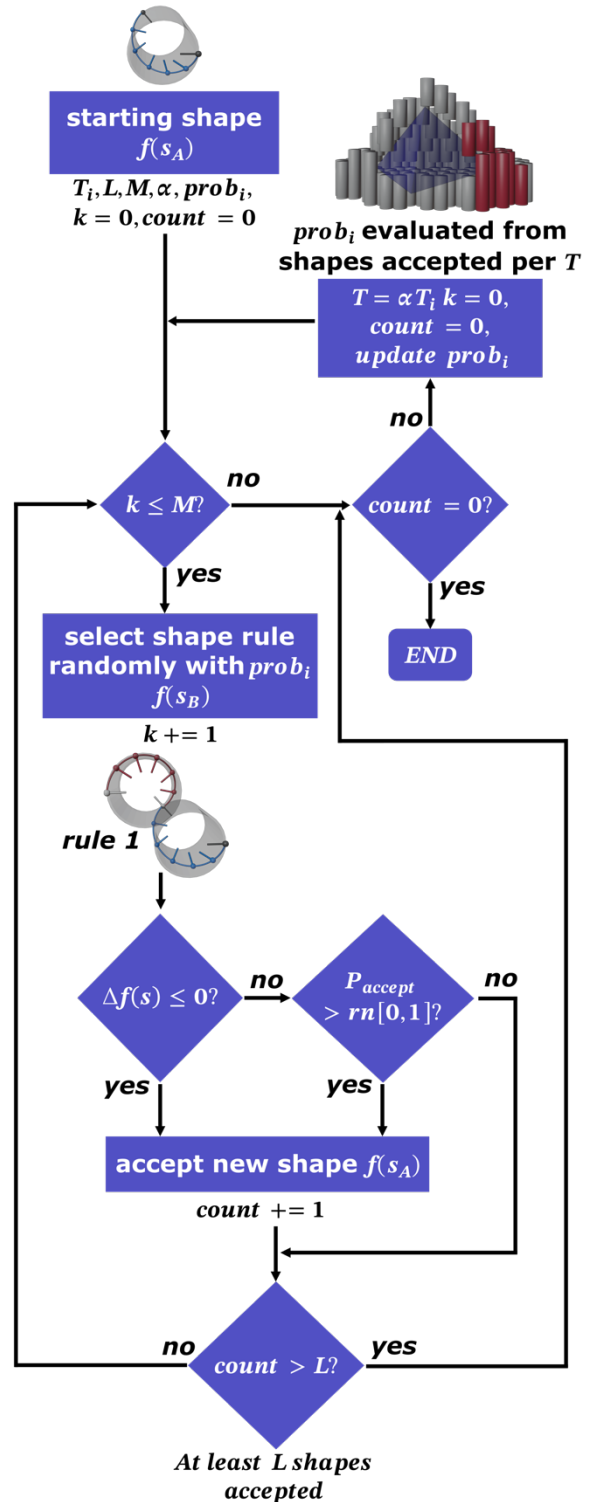
FIGURE 6: SHAPE ANNEALING ALGORITHM FLOWCHART

RMSF is one of the common structural analyses that measures the fluctuation of the DNA origami nanostructure during simulation. One way to calculate RMSF is by averaging the configurations generated over the entire simulation trajectory

[38]. It is important to note that all simulations in this work are performed at 295 K and a 15.15 fs time step (0.005 simulation units). All additional parameters used are detailed in literature by Doye et al. [31].

(a) $f(x_0) = 1.19^{x_0}$

(b) $f(x_1) = \frac{1}{(x_1 - 9.0)^6}$

(c) Objective function comparison
$f(x_1) = \frac{1}{(x_1 - 9.0)^6}$
$f(x_{fit}) = 2.04e\text{-}30\,e^{8.77 x_{fit}}$

(d)
*x_1 is distance to mesh wall
mesh envelope distance
(set desired distance)

FIGURE 7: $F(X)$ REARTICULATION (A) PREVIOUS $F(X_0)$ (B) CURRENT $F(X_1)$ (C) EXPONENTIAL VS. HYPERBOLIC $F(X)$ COMPARISON (D) ILLUSTRATION OF $F(X)$ VARIABLES.

2.5 Rearticulation of Coating Behavior

As stated previously, the objective function defines the desired characteristic of coating and is used to assess whether a shape rule is accepted or rejected. In previous work, coating behavior is achieved by minimizing the distance from the center of generated helical 7 bp scaffold sections towards the outer wall of the input triangular mesh. The distance is then scaled along the following objective function to articulate coating [22]:

$$f(x) = 1.19^x , \qquad (8)$$

where x is the minimum perpendicular Euclidean distance from the outer wall of the input mesh to the 7 bp scaffold section. The constant value in the previous objective function is purely arbitrary and is selected from a range of 1.00 to 2.00 where values approaching 1.00 generates designs that are loosely coated with higher coverage while values approaching 2.00 generates designs that are tightly coated with limited coverage of the polyhedral mesh. The rationale for selecting an exponential function is because the shape rules generated at a closer distance to the mesh have lower $f(x)$ values and therefore minimize $\Delta f(x)$ while the shape rules generated at a further distance from the mesh would have a higher likelihood of rejection due to much higher $f(x)$ values. In the previous technique, internal porosity or quality is traded for input mesh coverage. This is shown in the plot from Figure 7(a). Although this objective function demonstrates the capability of coating, the coated designs do not replicate the overall shape of the input polyhedral mesh as shown in Figure 4. This is due to the slow ramp of the exponential curve in Figure 7(a) where the shape rules generated at a closer distance to the mesh surface do not fully minimize $\Delta f(x)$ and the shape rules generated at a further distance from the mesh do not have comparably higher $f(x)$ values. Evidently, a change in objective function is necessary.

To increase quality and internal porosity of overall coated designs, the generated scaffold sections are limited to set desired distances measured from the outer wall of the desired mesh to a location on the outside of the mesh and the objective function is replaced with a hyperbolic function following a more step-like behavior:

$$f(x) = \frac{1}{(d-x)^6}, \qquad (9)$$

where x is the minimum perpendicular Euclidean distance from the outer wall of the input mesh to the 7 bp scaffold section and d is the set desired distance from outer wall of input mesh in which scaffold sections can be accepted to control effective wall thickness. The step-like behavior allows for more acceptance of feasible scaffold sections that tightly coat the outer wall of the input mesh without compromising porosity since the $f(x)$ values are all close to zero before the set desired distance. As the hyperbolic function approaches the set-desired distance, there is an abrupt increase in $f(x)$ values, which lowers the likelihood of shape rule acceptance according to the shape annealing algorithm. This is shown in Figure 7(b). This hyperbolic function also allows for physically relevant parameters as opposed to the arbitrary parameters of the exponential function shown in Figure 7(c). The exponent on the refined objective function is selected from a range of 2 to 9 where values approaching 2 show a more gradual increase in function values with distances close to d while values approaching 9 show a sharper increase in function values that become too computationally intensive with distances closer to d. 6 is a more appropriate exponent value since it

creates function values with distances closer to d that are computationally feasible to calculate and still maintain a sharp increase in function values. Figure 7(d) paints a picture of a 7 bp scaffold section accepted within the set desired distance.

2.6 Hustin Move Set Initial Probability

The Hustin move set initial probability used for all designs generated with the refined shape annealing algorithm in order of shape rules are 0.18 for the crossover rule, 0.44 for the extension rule, and 0.38 for the reversal rule. An initial probability search was implemented before changing the objective function with set desired distances by varying each shape rule from (0,1) to find structures that have the largest number of scaffold sections within an arbitrary distance of 12.0 nm from the mesh surface. This arbitrary distance was selected because the resulting scaffold routes showed sparse searches of the vast design space beyond $d = 12.0$ nm. The rationale for setting a bound for measuring the total scaffold sections is to find the shape rule probabilities with a higher porous layer close to the mesh surface. Since regions with large gaps from the previous algorithm occur due to a reduction in the application of the extension rule, the refined algorithm has a lower probability for the crossover rule and the highest probability for the extension rule. The reversal rule has a reasonably high probability to ensure that the algorithm does not converge on a local minimum, or the scaffold is not stuck during an anneal and can fully explore the design space. The probabilities are then dynamically adjusted during run time as described in Section 2.2.1.

3. RESULTS

By using a 3D tetrahedral mesh as input, the refined shape annealing algorithm can generate coated-type designs with a variety of effective wall thicknesses. After generating DNA origami designs of varying thicknesses, the designs are converted to the caDNAno JSON file format using a customized scadnano Python script [39]. The designs are then auto-stapled using the caDNAno auto-stapling function [2] with additional manual editing due to the non-traditional scaffold routings for effective stapling. After stapling, the JSON files are converted using the TacoxDNA platform [40] to PDB files, which are visualized using ChimeraX [41,42]. The JSON files are also converted to oxDNA topology and configuration files which are used as input files to run simulations. The output files after running simulations are used as input for the RMSF analysis scripts. The results from the analysis are then visualized using oxView [38] and ChimeraX.

Since the refined shape annealing algorithm uses a probabilistic technique, the generated continuous scaffold can get stuck, because the algorithm converges to a local minimum. This is typically avoided by selecting the best generated solution from several algorithm runs [43]. In this work, the algorithm is run in 10 batches of 10 (100 times in total), where the best generated solution is selected per batch. This is done per d from 8.0 to 12.0 nm. Since the metric of success is a tightly packed generation of scaffold within the set desired distance, the 10

DNA origami designs with the highest scaffold length are selected. An exhaustive search is performed to select the best shape annealing parameters per d that follow the metric of success stated by varying the limit (L) and temperature (T), which were found to be more sensitive parameters during an anneal. Table 1 shows the shape annealing parameters per set desired distance using the tetrahedral mesh as input. Table 2 shows the mean and standard deviation of 10 best generated solutions per 10 batches using the tetrahedral input mesh. To demonstrate the capability of the refined shape annealing algorithm in generating DNA origami designs with varying effective wall thicknesses, Figure 8 shows the cross-sectional view of the top three out of 10 designs with $d = [8.0, 9.0, 12.0]$ nm. Figure 9 shows the fully relaxed configurations of the structures generated in Figure 8 after structural validation through oxDNA simulations. Figure 10 shows the root mean-square fluctuations (RMSF) of the structures generated from oxDNA. Figure 12 illustrates the average effective wall thickness analysis of the top 10 generated DNA origami designs per batches of 10.

To ensure the method of analysis in Figure 12 is not mesh specific, the results from a cube input mesh with an edge length of 21.2 nm using the refined technique with $d = 9.0$ nm are also analyzed under the same method. Figure 11 shows an example of a fully relaxed configuration and RMSF from the 10 best generated DNA origami designs using the cube input mesh. In Figure 13, the same method of analysis of the top 10 generated DNA origami designs per batches of 10 for the cube input mesh is used to illustrate effective wall thickness. Tables 3 & 4 show the shape annealing parameters for $d = 9.0$ nm and the mean and standard deviation of the 10 best generated solutions per batch, respectively, using the input cube mesh.
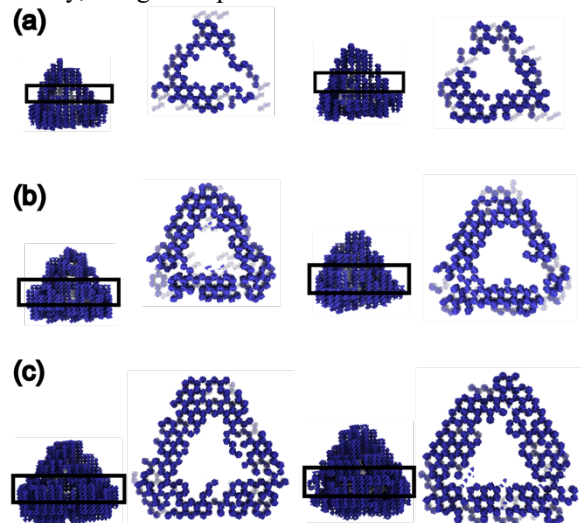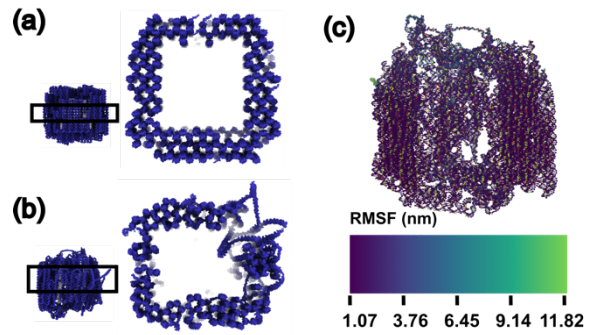

FIGURE 8: TWO OF BEST GENERATED DNA ORIGAMI DESIGNS FOR TETRAHEDRAL MESH USING REFINED TECHNIQUE WITH TOTAL SCAFFOLD LENGTHS OF: (A) 5,306 BP AND 5,257 BP FOR $D = 8.0$ NM, (B) 7,273 BP AND 6,979 BP FOR $D = 9.0$ NM, AND (C) 12,264 BP AND 12,586 BP FOR $D = 12.0$ NM.
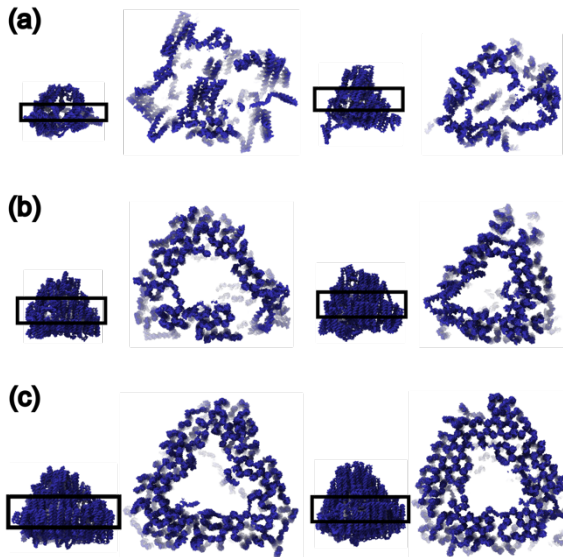
FIGURE 9: FULLY RELAXED CONFIGRUATIONS FROM OXDNA SIMULATIONS OF STRUCTURES IN FIGURE 8 FOR: (A) D = 8.0 NM, (B) D = 9.0 NM, AND (C) D = 12.0 NM.
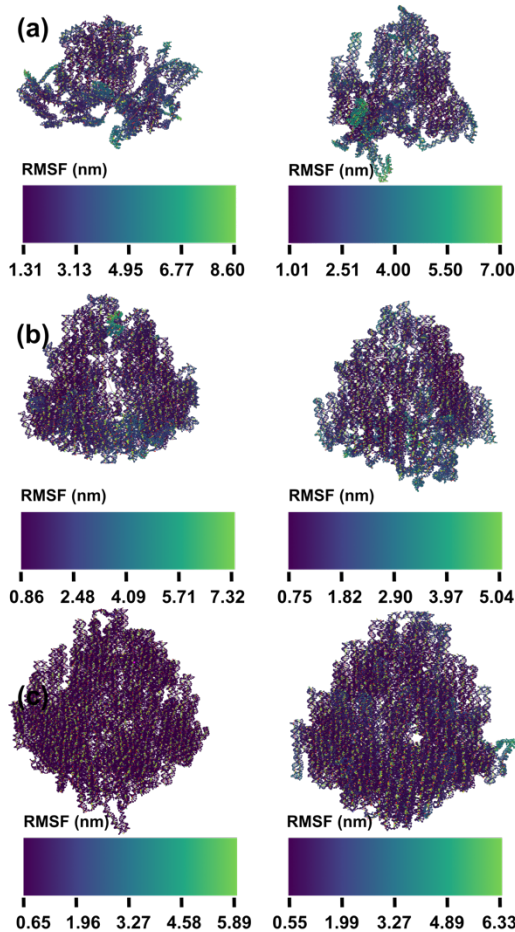
FIGURE 10: RMSF OF STRUCTURES IN FIGURE 8 FOR: (A) D = 8.0 NM, (B) D = 9.0 NM, AND (C) D = 12.0 NM. RMSF PATTERNS USE A COLORMAP FROM VIOLET TO YELLOW.


FIGURE 11: ONE OF BEST GENERATED DNA ORIGAMI DESIGNS USING REFINED TECHNIQUE AND CUBE MESH FOR D = 9.0 NM (A) WITH TOTAL SCAFFOLD LENGTH OF 13,937 NM (B) FULLY RELAXED CONFIGURATION (C) RMSF PATTERN USING COLORMAP FROM VIOLET TO YELLOW.
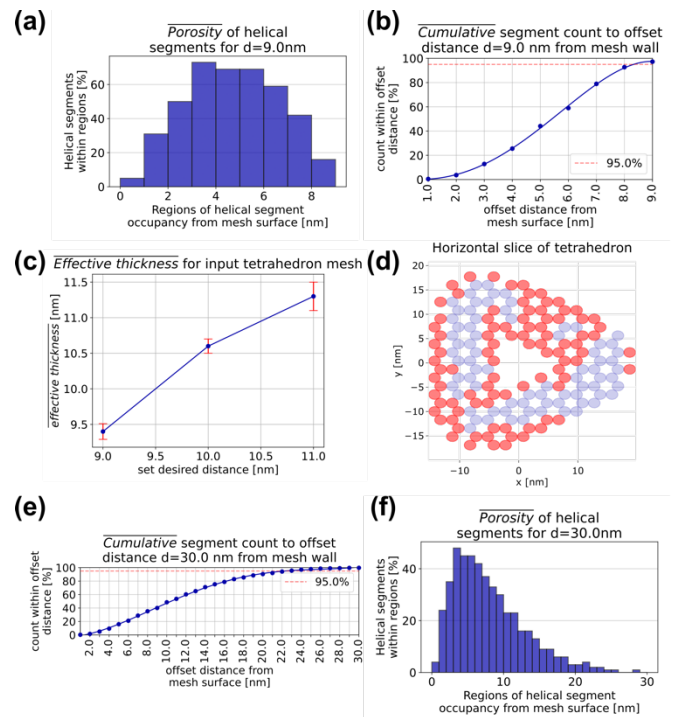
FIGURE 12: EFFECTIVE THICKNESS ANALYSIS FOR TETRAHEDRAL MESH USING REFINED TECHNIQUE: (A) AVG POROSITY ANALYSIS FOR D = 9.0 NM (B) AVG EFFECTIVE THICKNESS ANALYSIS FOR D = 9.0 NM (C) AVG EFFECTIVE THICKNESS COMPARISON (D) GRID FILLING VISUALIZATION AT MID-HEIGHT OF A STRUCTURE SHOWING UNOCCUPIED GRIDS IN RED. USING SIMPLE TECHNIQUE: (E) AVG EFFECTIVE THICKNESS ANALYSIS (F) AVG POROSITY ANALYSIS
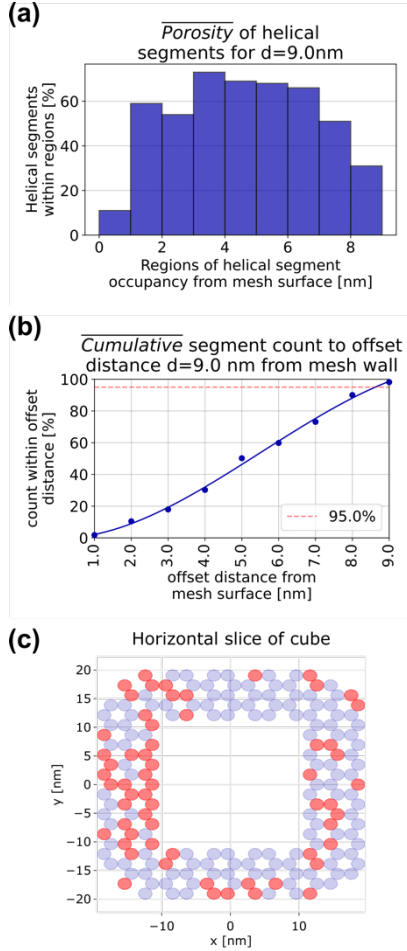
FIGURE 13: EFFECTIVE THICKNESS ANALYSIS FOR CUBE MESH USING REFINED TECHNIQUE FOR $D = 9.0$ NM (A) AVG POROSITY ANALYSIS (B) AVG EFFECTIVE WALL THICKNESS ANALYSIS (C) GRID FILLING VISUALIZATION AT MID-HEIGHT OF A STRUCTURE SHOWING UNOCCUPIED GRIDS IN RED.

TABLE 1 REFINED SHAPE ANNEALING OPTIMIZATION PARAMETERS USING TETRAHEDRAL MESH AS INPUT

d [nm]	T	L	α	M
8.0	956	280	0.98	510
9.0	941	300	0.98	510
10.0	980	260	0.98	510
11.0	888	340	0.98	510
12.0	912	260	0.98	510

TABLE 2 AVG AND STD DEV OF TOTAL SCAFFOLD LENGTH (sl) OF TOP 10 GENERATED DESIGNS PER 10 BATCHES WITH INPUT TETRAHEDRAL MESH USING REFINED TECHNIQUE

d [nm]	$\overline{sl}$	σ
8.0	5187	30
9.0	6755	38
10.0	8456	34
11.0	10549	69
12.0	12138	40

TABLE 3 REFINED SHAPE ANNEALING OPTIMIZATION PARAMETERS USING CUBE MESH AS INPUT

d [nm]	T	L	α	M
9.0	975	360	0.98	510

TABLE 4 AVG AND STD DEV OF TOTAL SCAFFOLD LENGTH (sl) OF TOP 10 GENERATED DESIGNS PER 10 BATCHES WITH INPUT CUBE MESH USING REFINED TECHNIQUE

d [nm]	$\overline{sl}$	σ
9.0	13818	49

4. DISCUSSION

Figure 8-10 illustrate the capability of the refined shape annealing algorithm in controlling the quality of the effective wall thickness of coated-type designs using a complex tetrahedral boundary. Figure 8 illustrates the tight coating of scaffold sections which is highlighted in the maintenance of the triangular shape toward the edges of the generated structures. This is unlike the structure generated in Figure 4 using the simple shape annealing algorithm where the outer walls of the structure form an arbitrary shape. Figures 9 and 10 illustrate the simulation results for the structures in Figure 8 with $d = [8.0, 9.0, 12.0]$ nm. Although the fully relaxed configurations and RMSF calculations in Figure 9 and 10 show major deviations for $d = 8.0$ nm from the intended final configuration, there are only slight deviations for $d = 9.0$ nm and higher. Figure 10 illustrates a decrease in regions with high fluctuations from increasing the set desired distance values. The results from the RMSF calculations in Figure 10 show some regions of high RMSF values where the crossover rule (rule 1) is accepted recursively with lack of variety in the acceptance of the crossover rule and extension rule (rule 2). These regions where rule 1 is recursively accepted are subjected to high fluctuations, which should be decreased before physical characterization. A post-processing step would be a prime solution to address such limitations by applying the new extension rule to those regions. Instead of scaffold sections that are only 7 bp in length, the scaffold sections would then be at least 14 bp in length which would create more effective stapling and reductions in the number of high fluctuations. In the future, it would be interesting to explore how the generated scaffold route is affected by the incorporation of RMSF data.

Figure 12 illustrates verification of increasing internal porosity and quality in changing effective wall thickness values of the generated scaffold sections of the top 10 designs. The distance of each helical segment in Figure 12 is calculated from its center. The set desired distances of 9.0 nm to 11.0 nm were selected for analysis since values lower than 9.0 nm are unable to generate structures maintaining structural integrity while values higher than 11.0 nm generate structures with scaffold lengths orders of magnitude longer than the M13 scaffold. Figure 12 (a) illustrates the average porosity within the walls of the top 10 generated designs using the refined technique by calculating the percent of helical segments generated out of the theoretical maximum helical segments within set regions from the mesh

9 © 2023 by ASME

surface to d. The plotted region starts from [0.0,1.0) nm and increases by 1.0 nm. Although Figure 12 (a) only shows the porosity analysis for $d = 9.0$ nm, the analyses for higher distances show similar trends. While Figure 12 (a) shows a trend that is approximately uniform with a higher percentage of helical segments within the set regions, Figure 12 (f) shows a trend that is approximately exponential with a lower percentage of helical segments within set regions which therefore demonstrates an increase in effective thickness quality and porosity. Improvement in uniformity can also be quantitatively measured with a coefficient of variation (CV) formula [44] which shows the diversity of a dataset with respect to the mean:

$$CV = \frac{\sigma}{\mu}, \tag{10}$$

where σ is the standard deviation and μ is the mean. The CV ratio is directly proportional to the variability within the data set in relation to the mean [44]. The histogram of the original algorithm in Figure 12 (f) has a CV of 0.96 while the histogram of the updated algorithm shows improvements with a CV of 0.49.

Using the refined technique, an effective wall thickness value is calculated from the cumulative percent of helical segments within an offset distance from the mesh surface to d in Figure 12 (b). The offset distance with 95.0% of total scaffold sections plus 1.0 nm to account for the radius of the scaffold is the effective wall thickness value. Although Figure 12(b) only shows the effective wall thickness analysis for $d = 9.0$ nm, the analyses for higher distances show similar trends. Figure 12 (c) shows an increase in effective wall thickness value that approximates to the respective set desired distance using the refined technique. Figure 12 (e), which uses the simple technique, has a slower ramp to the total number of scaffold sections than in Figure 12 (b), demonstrating a steady decrease in generated scaffold segments towards to the edges of the structure. This steady decrease demonstrates the lack of generated structures with tight coatings unlike the plot illustrated in Figure 12 (b). Figure 12 (d) provides a visualization of the filling within the walls at the middle of the last generated structure with $d = 9.0$ nm from Figure 8 (a) using the refined technique with 45% of grids filled, which certainly shows a tighter packing than the previous technique in Figure 4.

Using the refined technique, a simple cube mesh is used as an input with $d = 9.0$ nm to demonstrate the generality of this analysis method. Figure 11 (a) illustrates the versatility of the refined shape annealing algorithm in utilizing arbitrary input mesh files to generate unique scaffold routes. Figure 11 (b) shows a slight deviation from the intended final configuration. The same conclusions can be made in Figure 10 for Figure 11 (c) where regions of high RMSF values are due to lack of variety in sequentially accepted shape rules. Using the same analysis method in Figure 12 (a), Figure 13(a) illustrates the average porosity with the cube as an input mesh that also follows an approximately uniform trend. However, Figure 13(a) has a CV of 0.36 thus showing more uniformity within the wall of the structure with a cube input mesh than a tetrahedron input mesh. Using the same analysis method in Figure 12 (b), the average effective thickness value calculated is 9.0 nm. Interestingly, in

Figure 13 (c) 69% of grids are filled, which is much higher than the number of grids filled with the tetrahedral mesh as input. This leads to the hypothesis that prism-like meshes have a higher porosity due to the alignment of the helical axis to the mesh wall. In the future, it would be interesting to investigate the relationship between porosity and scaffold axis orientation.

Although the refined algorithm shows major improvements in the articulation of coating, Figure 8 and 11 show small gaps within the walls of the generated designs. This is attributed to the scaffold being unable to fully search the complex design space during an anneal. A post-processing step can also be used to address the lack of total design space exploration in the future. Expanding our simple shape grammar rules would be an interesting way of solving the gap problem. This step will only target regions with small gaps to ensure adequate filling. The same shape annealing algorithm would run, with only one differing property where a new shape grammar rule would replace the current extension rule (rule 2) where the scaffold could be extended by 14 bp instead of 7 bp.

In addition to the application of a post-processing step, future work will focus on automating the stapling process for full design automation. Another desired next step would be developing an algorithm to unite the start and end location of the generated scaffold pattern to accommodate both linear and circular scaffolds.

5. CONCLUSION

The results show that shape annealing is apt for automating varying effective wall thicknesses to generate novel DNA origami designs with high dense layering within the walls. By applying design methods, the refined shape annealing algorithm can navigate the challenging physical space. Such design methods include adding the Hustin move set to favor rules that follow the objective function and redefining coating to show that desired characteristics can be well coded into DNA origami designs. The results, which introduce new capabilities to enable the generation of an array of effective quality solutions, are an important step for creating a tool for the DNA origami design community. This opens doors to many desirable traits to address the long-standing challenges in DNA origami design that could be coded into the structures such as increasing nanostructure yield [45] and closing the design-iteration loop due to added knowledge of design-property relationship [46].

REFERENCES

[1] Williams, S., Lund, K., Lin, C., Wonka, P., Lindsay, S., and Yan, H., 2009, "Tiamat: A Three-Dimensional Editing Tool for Complex DNA Structures," *DNA Computing*, A. Goel, F.C. Simmel, and P. Sosík, eds., Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 90–101.

[2] Douglas, S. M., Marblestone, A. H., Teerapittayanon, S., Vazquez, A., Church, G. M., and Shih, W. M., 2009, "Rapid Prototyping of 3D DNA-Origami Shapes with CaDNAno," Nucleic Acids Res, **37**(15), pp. 5001–5006.

[3] Jun, H., Zhang, F., Shepherd, T., Ratanalert, S., Qi, X., Yan, H., and Bathe, M., 2019, "Autonomously Designed Free-Form 2D DNA Origami," Science Advances, **5**(1), p. eaav0655.

[4] Huang, C.-M., Kucinic, A., Johnson, J. A., Su, H.-J., and Castro, C. E., 2020, *Integrating Computer-Aided Engineering and Computer-Aided Design for DNA Assemblies*, Bioengineering.

[5] Benson, E., Mohammed, A., Gardell, J., Masich, S., Czeizler, E., Orponen, P., and Högberg, B., 2015, "DNA Rendering of Polyhedral Meshes at the Nanoscale," Nature, **523**(7561), pp. 441–444.

[6] Veneziano, R., Ratanalert, S., Zhang, K., Zhang, F., Yan, H., Chiu, W., and Bathe, M., 2016, "Designer Nanoscale DNA Assemblies Programmed from the Top Down," Science, **352**(6293), pp. 1534–1534.

[7] Jun, H., Shepherd, T. R., Zhang, K., Bricker, W. P., Li, S., Chiu, W., and Bathe, M., 2019, "Automated Sequence Design of 3D Polyhedral Wireframe DNA Origami with Honeycomb Edges," ACS Nano, **13**(2), pp. 2083–2093.

[8] Jun, H., Wang, X., Bricker, W. P., Jackson, S., and Bathe, M., 2020, "Rapid Prototyping of Wireframe Scaffolded DNA Origami Using ATHENA," p. 2020.02.09.940320.

[9] 2021, "Lcbb/Athena."

[10] "METIS – Three-Dimensional, Algorithmically-Generated Library of DNA Origami Shapes" [Online]. Available: https://metis-dna-origami.org/. [Accessed: 06-Feb-2022].

[11] Seeman, N. C., 1982, "Nucleic Acid Junctions and Lattices," Journal of Theoretical Biology, **99**(2), pp. 237–247.

[12] Rothemund, P. W. K., 2006, "Folding DNA to Create Nanoscale Shapes and Patterns," Nature, **440**(7082), pp. 297–302.

[13] Seeman, N. C., and Sleiman, H. F., 2017, "DNA Nanotechnology," Nature Reviews Materials, **3**(1), pp. 1–23.

[14] Tørring, T., and Gothelf, K. V., 2013, "DNA Nanotechnology: A Curiosity or a Promising Technology?," F1000Prime Rep, **5**.

[15] Castro, C. E., Kilchherr, F., Kim, D.-N., Shiao, E. L., Wauer, T., Wortmann, P., Bathe, M., and Dietz, H., 2011, "A Primer to Scaffolded DNA Origami," Nat Methods, **8**(3), pp. 221–229.

[16] Said, H., Schüller, V. J., Eber, F. J., Wege, C., Liedl, T., and Richert, C., 2012, "M1.3 – a Small Scaffold for DNA Origami," Nanoscale, **5**(1), pp. 284–290.

[17] Sun, W., Boulais, E., Hakobyan, Y., Wang, W. L., Guan, A., Bathe, M., and Yin, P., 2014, "Casting Inorganic Structures with DNA Molds," Science, **346**(6210), pp. 1258361–1258361.

[18] Du, K., Park, M., Ding, J., Hu, H., and Zhang, Z., 2017, "Sub-10 Nm Patterning with DNA Nanostructures: A Short Perspective," Nanotechnology, **28**(44), p. 442501.

[19] Douglas, S. M., Bachelet, I., and Church, G. M., 2012, "A Logic-Gated Nanorobot for Targeted Transport of Molecular Payloads," Science, **335**(6070), pp. 831–834.

[20] Selnihhin, D., Sparvath, S. M., Preus, S., Birkedal, V., and Andersen, E. S., 2018, "Multifluorophore DNA Origami Beacon as a Biosensing Platform," ACS Nano, **12**(6), pp. 5699–5708.

[21] Cagan, J., and Mitchell, W. J., 1993, "Optimally Directed Shape Generation by Shape Annealing," Environ. Plann. B, **20**(1), pp. 5–12.

[22] Babatunde, B., Arias, D. S., Cagan, J., and Taylor, R. E., 2021, "Generating DNA Origami Nanostructures through Shape Annealing," Applied Sciences, **11**(7), p. 2950.

[23] Stiny, G., 1980, "Introduction to Shape and Shape Grammars," Environ. Plann. B, **7**(3), pp. 343–351.

[24] Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P., 1983, "Optimization by Simulated Annealing," Science, **220**(4598), pp. 671–680.

[25] Hustin, S., and Sangiovanni-Vincentelli, A., 1987, "TIM, a New Standard Cell Placement Program Based on the Simulated Annealing Algorithm," Hilton Head, SC.

[26] Szykman, S., and Cagan, J., 1996, "Synthesis of Optimal Nonorthogonal Routes," Journal of Mechanical Design, **118**(3), pp. 419–424.

[27] Shea, K., and Cagan, J., 1999, "Languages and Semantics of Grammatical Discrete Structures," AIEDAM, **13**(4), pp. 241–251.

[28] Kim, D.-N., Kilchherr, F., Dietz, H., and Bathe, M., 2012, "Quantitative Prediction of 3D Solution Shape and Flexibility of Nucleic Acid Nanostructures," Nucleic Acids Res, **40**(7), pp. 2862–2868.

[29] Maffeo, C., and Aksimentiev, A., 2020, "MrDNA: A Multi-Resolution Model for Predicting the Structure and Dynamics of DNA Systems," Nucleic Acids Research, **48**(9), pp. 5135–5146.

[30] Yoo, J., and Aksimentiev, A., 2013, "In Situ Structure and Dynamics of DNA Origami Determined through Molecular Dynamics Simulations," Proceedings of the National Academy of Sciences of the United States of America, **110**(50), pp. 20099–20104.

[31] Doye, J. P. K., Fowler, H., Prešern, D., Bohlin, J., Rovigatti, L., Romano, F., Šulc, P., Wong, C. K., Louis, A. A., Schreck, J. S., Engel, M. C., Matthies, M., Benson, E., Poppleton, E., and Snodin, B. E. K., 2020,

"The OxDNA Coarse-Grained Model as a Tool to Simulate DNA Origami," arXiv:2004.05052 [cond-mat].

[32] Ouldridge, T. E., Louis, A. A., and Doye, J. P. K., 2011, "Structural, Mechanical, and Thermodynamic Properties of a Coarse-Grained DNA Model," J. Chem. Phys., **134**(8), p. 085101.

[33] Šulc, P., Romano, F., Ouldridge, T. E., Rovigatti, L., Doye, J. P. K., and Louis, A. A., 2012, "Sequence-Dependent Thermodynamics of a Coarse-Grained DNA Model," The Journal of Chemical Physics, **137**(13), p. 135101.

[34] Snodin, B. E. K., Randisi, F., Mosayebi, M., Sulc, P., Schreck, J. S., Romano, F., Ouldridge, T. E., Tsukanov, R., Nir, E., Louis, A. A., and Doye, J. P. K., 2015, "Introducing Improved Structural Properties and Salt Dependence into a Coarse-Grained Model of DNA," The Journal of Chemical Physics, **142**(23), p. 234901.

[35] Snodin, B. E. K., Schreck, J. S., Romano, F., Louis, A. A., and Doye, J. P. K., 2019, "Coarse-Grained Modelling of the Structural Properties of DNA Origami," Nucleic Acids Research, **47**(3), pp. 1585–1597.

[36] Engel, M. C., Smith, D. M., Jobst, M. A., Sajfutdinow, M., Liedl, T., Romano, F., Rovigatti, L., Louis, A. A., and Doye, J. P. K., 2018, "Force-Induced Unravelling of DNA Origami," ACS Nano, **12**(7), pp. 6734–6747.

[37] Shi, Z., and Arya, G., 2020, "Free Energy Landscape of Salt-Actuated Reconfigurable DNA Nanodevices," Nucleic Acids Research, **48**(2), pp. 548–560.

[38] Poppleton, E., Bohlin, J., Matthies, M., Sharma, S., Zhang, F., and Šulc, P., 2020, "Design, Optimization and Analysis of Large DNA and RNA Nanostructures through Interactive Visualization, Editing and Molecular Simulation," Nucleic Acids Res, **48**(12), p. e72.

[39] Doty, D., Lee, B. L., and Stérin, T., 2020, "Scadnano: A Browser-Based, Scriptable Tool for Designing DNA Nanostructures," arXiv:2005.11841 [cs, q-bio].

[40] Suma, A., Poppleton, E., Matthies, M., Šulc, P., Romano, F., Louis, A. A., Doye, J. P. K., Micheletti, C., and Rovigatti, L., 2019, "TacoxDNA: A User-Friendly Web Server for Simulations of Complex DNA Structures, from Single Strands to Origami," Journal of Computational Chemistry, **40**(29), pp. 2586–2595.

[41] Goddard, T. D., Huang, C. C., Meng, E. C., Pettersen, E. F., Couch, G. S., Morris, J. H., and Ferrin, T. E., 2018, "UCSF ChimeraX: Meeting Modern Challenges in Visualization and Analysis," Protein Science, **27**(1), pp. 14–25.

[42] Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., Morris, J. H., and Ferrin, T. E., 2021, "UCSF ChimeraX: Structure Visualization for Researchers, Educators, and Developers," Protein Science, **30**(1), pp. 70–82.

[43] Rutenbar, R. A., 1989, "Simulated Annealing Algorithms: An Overview," IEEE Circuits and Devices Magazine, **5**(1), pp. 19–26.

[44] Everitt, B. S., 2002, *The Cambridge Dictionary of Statistics*, Cambridge University Press, Cambridge, UNITED KINGDOM.

[45] Ke, Y., Bellot, G., Voigt, N. V., Fradkov, E., and Shih, W. M., 2012, "Two Design Strategies for Enhancement of Multilayer–DNA-Origami Folding: Underwinding for Specific Intercalator Rescue and Staple-Break Positioning," Chem. Sci., **3**(8), pp. 2587–2597.

[46] Majikes, J. M., and Liddle, J. A., 2020, "DNA Origami Design: A How-To Tutorial," J. RES. NATL. INST. STAN., **126**, p. 126001.