PHYSIK-DEPARTMENT

Computational Analysis of DNA Origami Nanostructures using Atomic Model Building

Dissertation by

Elija Feigl

TECHNISCHE UNIVERSITÄT MÜNCHEN

TUM

Technische Universität München

TUM School of Natural Sciences

Computational Analysis of DNA Origami Nanostructures using
Atomic Model Building

Elija Feigl

List of Peer-Reviewed Publications

Feigl E.*, Kube M.*, Kohler F.*, Nagel-Yüksel N., Willner E.M., Funke J.J., Gerling T., Stömmer P., Martin T.G., Scheres S.H.W., Dietz H.°
Revealing the structures of megadalton-scale DNA complexes with nucleotide resolution
Nature Communications 11 (2020)

Bertosin E.*, Stömmer P.*, Feigl E., Wenig M., Honemann M.N., Dietz H.°
Cryo-Electron Microscopy and Mass Analysis of Oligolysine-Coated DNA Nanostructures
ACS Nano 15, 6 (2021)

Hayakawa D., Videbæk T.E., Hall D.M., Fang H., Sigl C., Feigl E., Dietz H., Fraden S., Hagan M.F., Grason G.M., Rogers W.B.°
Geometrically programmed self-limited assembly of tubules using DNA origami colloids
PNAS 119, 43 (2022)

(*= These authors contributed equally to this work °= Corresponding author.)

Abstract

Deoxyribonucleic acid nanotechnology utilizes programmable self-assembly for the rational design of custom-shaped and functional macromolecules. The compact lattice-based structures designed with DNA origami produce micrometer-sized objects of defined three-dimensional shape on the megadalton-scale, by weaving a long DNA template strand together by using hundreds of shorter unique oligonucleotides. These DNA objects are investigated and experimentally validated in solution using cryogenic electron microscopy, generating volumetric electron-density data. Here, we address the problem of accurate structural interpretation of cryogenic electron microscopy maps of DNA nanostructures with computational data analysis. By developing semi-automated protocols for molecular dynamics-based pseudo-atomic model building, we enable annotation and contextualization of the electron density. The established methodology is distributed as an open-source software bundle that implements supplementary analysis tools. Among other features, our pseudo-atomic models enable targeted zoning and inspection of isolated design motifs, structural comparison of design iterations, and reproducible quantification of global object properties. In addition to facilitating precision analysis of experimental data on DNA origami objects, we show that the published pseudo-atomic models provide a reference resource for the validation of structure prediction models. This increase in available pseudo-atomic model coordinates allows designers to perform iterative base-pair level design improvement, creating an interdisciplinary positive feedback loop with computational structure prediction and modeling.

Contents

Introduction

Scaffolded DNA origami [1] enables scientists to design custom-shaped nanostructures up to the gigadalton scale [2]. Each of the several thousand nucleotides comprising such structures can be addressed and modified individually. Their placement in the three-dimensional assembly is encoded by the sequence of the DNA strand and facilitates the placement of building blocks with nanometer precision. In recent years, the field of DNA nanotechnology has achieved a significant increase in structural complexity. Also, efficient experimental protocols for structure validation have been established. These efforts created an ever-growing pool of available motifs and functionalities. Most current designs are multilayered, lattice-based, and enhanced with functional motifs and DNA handles to attach to other macromolecules [3, 4]. These molecules can be placed with Bohr radius resolution at specific nucleotides [3–5]. Additionally, lattice types of varying compactness are available, each with distinct benefits and applications. They range from dense, square-lattice layered DNA origami to the mostly hollow but extremely versatile Wire-frame [6–8] designs. Design motifs include hinges [9] and rotary axes [10, 11], control over curvature and twist [12], shape-complementary connection contacts [10, 13], as well as angle- and force- measuring features [5, 14]. Advancements in the field also include object stabilization by coating [15, 16] the DNA nanostructure and by chemical modifications [17]. These treatments inhibit denaturation at low salt concentration and nuclease degradation.

Such DNA nanostructures are used in a broad spectrum of research fields, including structural biology [10, 16, 18–20], therapeutics [3, 13, 21, 22], and plasmonics [23–25]. As the desired functionalities strongly depend on the specific shape and function of the DNA origami nano-object, increasing the stability and precision of these assemblies remains an important objective.

However, the design process of the nanostructures' overall shape still relies on abstraction and simplification of the structures' core components, B-DNA duplexes [26, 27]. Most of our current knowledge about the structure of helical DNA is based on diffraction data generated from crystalline DNA [28, 29]. In the context of DNA origami, the flexibility of

these building blocks is represented by the bending rigidity [30] of B-DNA duplexes. Yet, for large DNA nanostructures in solution deviations from the crystal structure are common due to the complex interplay of the lattice topology and the individual double-stranded segments. On a larger scale, for example, the crossover motif deviates from the naturally occurring Holliday junction configuration as the arms are aligned within the lattice. On the nucleotide level, the helical rise and twist are dependent on the sequence, affecting the overall duplex shape. The impact of these irregularities on megadalton DNA objects is mostly underrepresented in present design concepts.

Acquisition of high-resolution structural data, in conjunction with the development of high fidelity computational structure prediction of DNA nanostructures, allows analysis on a level previously unreachable [31–33]. Nucleotide level accuracy in the verification of nano-objects facilitates the high-precision iterative design of these structures. Until recently, only little high-resolution electron density data has been available [33], as most cryogenic electron microscopy volumetric data of DNA nanostructures is reported at comparatively low resolutions. As part of a cooperative effort comprising this work, we recently published 26 DNA origami structures [32] including electron density data with sub 5 Å local resolution.

Availability of cryogenic electron microscopy, single-particle analysis for objects at resolutions below 10 Å provides insights into the global shape of nanoscale objects [33]. However, the driving influences on structural flexibility and deviation from the target shape require the addressability of individual nucleotides. Neither nucleobase sequence, nor the topology of the macromolecule or its nucleotide identity with respect to the design, are represented by the electron-density data. In order to quantify the geometric properties of the nano-object, localization of each of its constituent building blocks within the volumetric is required. A solution widely applied for proteins is the construction of pseudo-atomic models fitted into electron density maps [34, 35]. These models present the structural context needed to interpret local features and provide a link from the intensity values, stored in each voxel of these datasets, to the geometry and topology of the object they represent.

A similar construction of pseudo-atomic models of DNA origami structures would enable the development of standardized analysis workflows and comparative structural studies of high-resolution experimental data. Yet, most DNA origami structures are too large for most available computational model building tools [36], due to the non-linearly increasing computational cost with the number of atoms and the size of nanostructures. The first pseudo-atomic model was published by Bai et al. in 2012, but only one structure [37] has been deposited since. In their protocol, designs had to be subdivided with each subset fitted individually and recombined manually, possibly impeding widespread utilization of the technique. A specialized tool or protocol for fast pseudo-atomic model building for

DNA origami nanostructures did not exist.

The lack of available pseudo-atomic models of DNA origami structures impacts a majority of computational models, as they are parametrized on limited data [26, 31, 38, 39]. This restriction drastically limits how the predictions made by these models generalize, reducing their usability. One way to solve this is to expand the pool of available structural information. Besides the lack of high-resolution data, the effort associated with the original protocol [33] likely hindered further developments in recent years. Here, we report a semi-automated protocol to create pseudo-atomic models for DNA origami designs within 24 hours of computing time on a standard desktop computer. The protocol is tailor-made for megadalton lattice-based DNA nanostructures and implemented as the Python-based application called *dnaFit* that includes customizable analysis workflows.

Research Target

In this work, we present a computational technique designed to complement experimental high-resolution cryogenic electron microscopy data of DNA origami nanostructures with the context and geometry provided by pseudo-atomic models. We developed a novel modeling protocol that accounts for the size of our nano-objects and exploits the particularities of compact lattice-based DNA origami. The generated pseudo-atomic structural data enriches the electron density data with the annotation of the atomic identities and the structure of their geometry. The application of this workflow does not require expert knowledge and is achievable with standard computational resources on the timescale of hours. The goal of this effort is to investigate and describe the global shape, flexibility, and inhomogeneity of DNA nanostructures, as well as to enable feature engineering and validation with nanometer accuracy. We discuss how the availability of pseudo-atomic models enables performing high-precision iterative design of 3D DNA origami structures on the sub-nanometer scale. Also, we introduce software tools that facilitate accurate analysis of the acquired experimental data. Finally, we use the coordinates from the pseudo-atomic model to validate and improve computational structure prediction models.

Structure of the Thesis

In Chapter 1, theoretical background on the basics of DNA origami nanotechnology and the structure of crystalline B-DNA is given. This chapter also includes an overview of pseudo-atomic model building and computational structure prediction of DNA nanostructures.

In Chapter 2, our protocol *Shrink-Wrap Fitting* for building pseudo-atomic models is introduced together with the DNA origami model building framework *dnaFit* that implements it.

In Chapter 3, we present the zoning and masking tool *FitViewer*, which uses structural annotation provided by the pseudo-atomic models to inspect subsets of electron density data.

In Chapter 4, several models generated using *Shrink-Wrap Fitting* are analyzed in more detail. The geometric information presented by atomic coordinates is used to quantify their geometry and aid in the iterative design process of DNA nanostructures.

In Chapter 5, we compare our high-resolution data to the latest structure prediction methods in the field of lattice-based DNA origami nanotechnology. The pseudo-atomic models from the previous chapters serve as validation for the accuracy of these methods.

In Chapter 6, this work concludes with the discussion of the impact of the availability of pseudo-atomic model construction on the field of DNA nanotechnology.

Chapter 1

Theoretical Background

Throughout this chapter, we describe the theoretical context of computational structure prediction and pseudo-atomic modeling of DNA origami nanostructures. We first present the discovery of the atomic structure of B-DNA (Section 1.1.1) and illustrate how its distinct topological constraints are utilized in the field of DNA origami nanotechnology (Section 1.1.1). Then, we introduce the imaging technique of single-particle cryogenic electron microscopy (cryo-EM) and three-dimensional structure reconstruction (Section 1.2). These methods generate the experimental data discussed throughout this work, both in terms of exploratory data analysis and validation of theoretical predictions.

Furthermore, we outline the concepts of molecular dynamics (MD) simulation in general (Section 1.3.1), with a particular focus on coarse graining and computational structure prediction of DNA origami nanostructures (Section 1.3.2). Lastly, we describe pseudo-atomic model building for volumetric electron density data of macromolecular structures centering on the technique of MD flexible fitting

1.1 DNA Nanotechnology

The B-DNA form exhibits a right-handed double-helical structure [40–43] produced by a quadripartite nucleotide pairing scheme [28, 44]. Under physiological conditions B-DNA is the prevalent form of DNA [45] and studied extensively for its role in encoding, replicating, and transmitting hereditary information [46, 47]. In recent years, the stability, programmability, bio-compatibility, and well-studied tertiary structure of B-DNA have promoted its repurposing as a construction material for nanotechnology [48]. Here, we present an atomic-level overview of the topology, structure, and dynamics of the B-DNA form required for structure validation, prediction, and coarse-grained modeling of DNA

nanostructures. We also provide a terse introduction of the technique of DNA origami used to construct the nano-objects presented throughout this work.

1.1.1 Atomistic Structure of DNA

In the decades after its discovery, with rapid advancements in imaging technology and crystallography, the atomic structure of double-helical B-DNA (Figure 1.1A) has been characterized thoroughly [29, 30, 49–51]. This dataset enables the description of a double-stranded helical DNA segment in terms of its overall properties like the helical pitch, diameter, and bending rigidity.
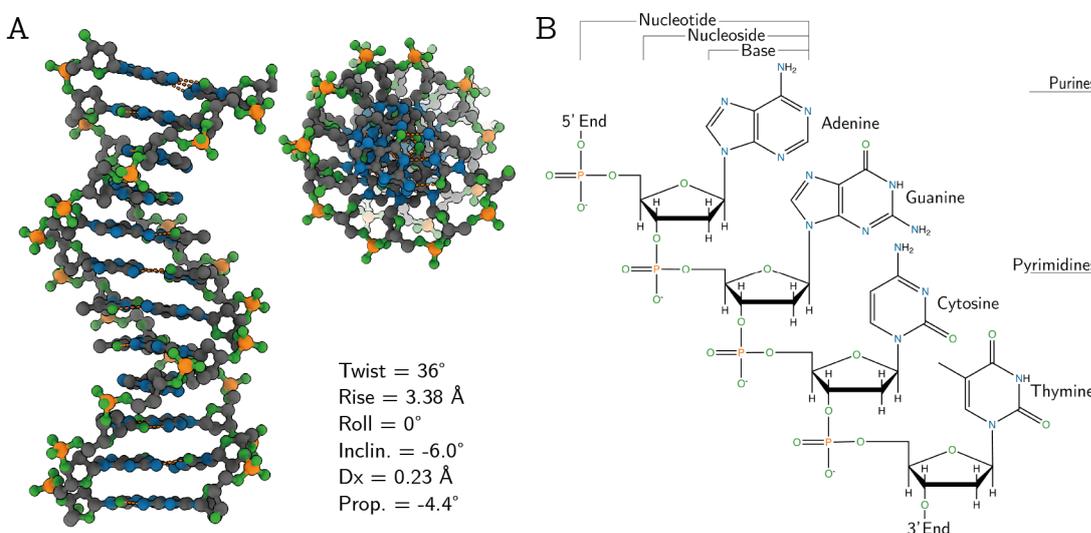
Figure 1.1 | **Atomic Structure and Topology of B-DNA: (A)** *Pseudo-atomic structure of 11 bases long B-DNA duplex from the side (left) and along the helical direction (right, top). The double-helix can be described in terms of its macroscopic parameters [52] (right, bottom).* **(B)** *Structural formula and nomenclature of a DNA strand of the four common bases adenine, guanine, thymine, and cytosine.*

In addition to this macroscopic perspective, the internal structure can be characterized at the atomic level (Figure 1.1B). The advent of computational physics and force field modeling [53–56] at the turn of the century and subsequent computational studies helped to complete the picture [57–60]. Nomenclature and definitions follow an established standard, referenced in the following from a review by Shing and Carter [52].

Each DNA strand is composed of a series of four common nucleobases (adenine, guanine, thymine, and cytosine) connected by the negatively charged sugar-phosphate backbone. In general atoms of the backbone are differentiated from the nucleobase by denotation with an apostrophe "prime" after the atom name, which follows the standard numbering scheme.

Numbering for the sugar starts with the carbon that connects nucleobase and backbone, while nucleobases start numbering at the nitrogen of the pyrimidine ring. Directionality of the strand can be determined by the carbon atoms of the backbone, with the standard being established as C5' to C3' "5prime-3prime".
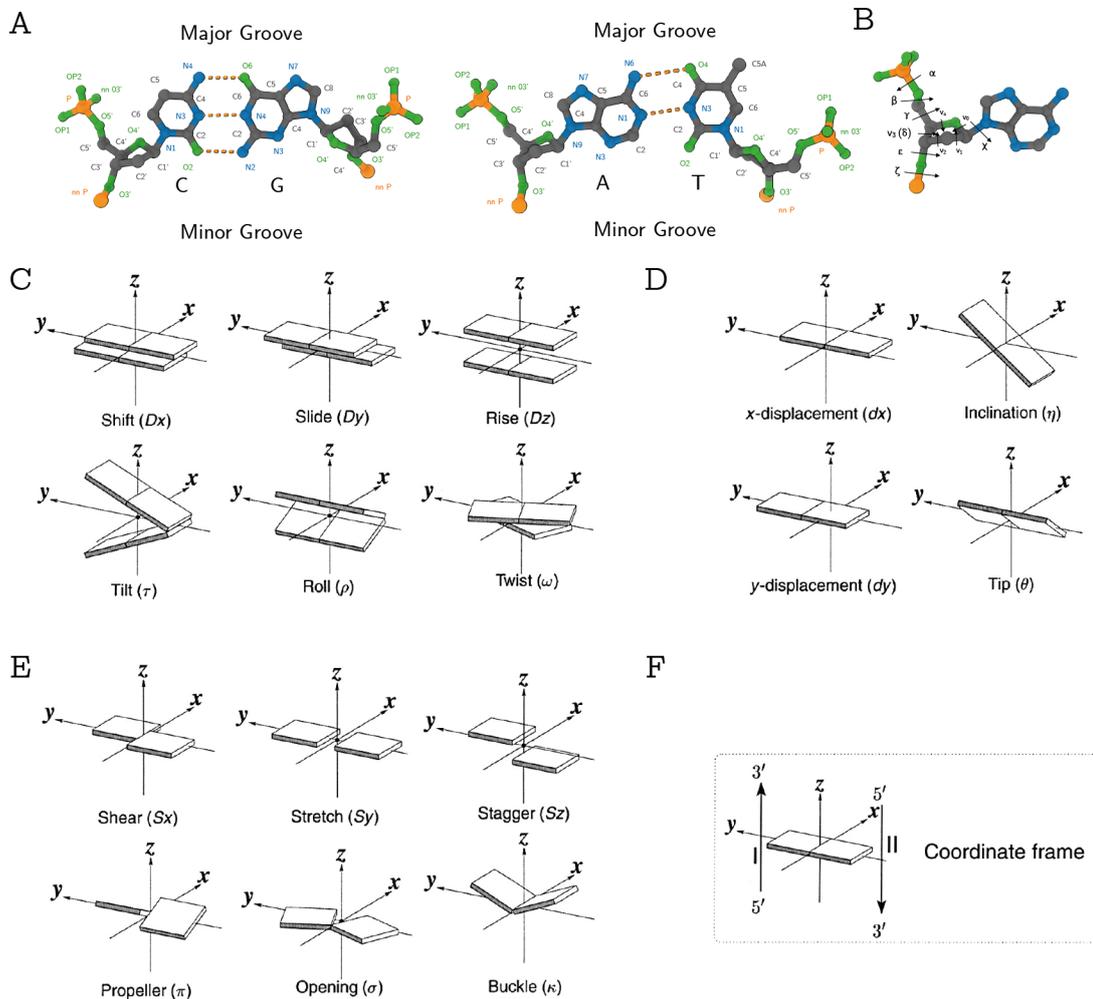
Figure 1.2 | *Nucleotide-Level Structure of B-DNA: (A) Structure formula including atomic nomenclature of the two canonical purine (R) — pyrimidine (Y) base-pairs. The pairs interact by forming up to three intermolecular hydrogen bonds. The orientation of the individual nucleotides determines the grooving of the helix, with the major groove located opposite the deoxyribose. **(B)** The conformation of the individual nucleotide is mostly determined by the six dihedral angles of the sugar-phosphate backbone and the sugar-base axis. **(C-F)** The structure of B-DNA is characterized by six base-pair (C), four base-pair helical (D), and six base-step (E) properties based on a standardized reference frame (F). Adopted from [61].*

While forming a double-stranded helix, the two strands interact via hydrogen bonding and $\pi - \pi$ orbital stacking of the nucleobases. The orbital stacking is most important for the stability of the helix [62, 63].

For B-DNA only two distinct combinations of one purine (R) and one pyrimidine (Y) are accessible, with the cytosine-guanine pair forming three hydrogen bonds and the adenine-thymine pair two (Figure 1.1A). The angle formed by these duplexes causes grooving of the helix with the minor groove being at the side of the C1' sugar carbon. With the nucleobases being planar, the nucleotide conformation is defined by six dihedral angels along the backbone and one at the attachment site of the nucleobase with the C1' atom of the deoxyribose (Figure 1.1B). Deviations from the planar arrangement of the nucleobase-pair are the second important parameter set that defines the helical conformation. These degrees of freedom have been combined into a distinct set of observables for both the base-pair and base-pair steps (Figure 1.1C). Originally perceived on a macroscopic level these properties are also defined in terms of standardized atomic coordinate reference frame [64, 65].

The study of the sequence dependence of these parameters [59, 66–72] and their correlation with deviations in the dihedral angels [58, 73, 74] remain active fields of research. The adequate modeling presents a challenge for atomic-level force fields [75] for MD simulations as well as coarse-grained modeling [39, 76].

1.1.2   Megadalton Lattice-Based DNA Origami

The programmability of reverse complementary base-pairing, combined with the stability and relative rigidity of the double helix provides the foundation for modern DNA nanotechnology. The flexibility of the DNA backbone facilitates the design of topological motifs that help connect double-stranded building blocks. Importantly, these connective patterns must not result in structural weak points. Inspired by nature, Ned Seeman pioneered the field in the 1980s when he repurposed the Holliday crossover motif [77] as a junction to connect multiple double-stranded DNA segments in two-dimensional arrays.[48, 78]. The helical segments, with a persistence length of 450 to 500 Å in a solution of medium to high ionic strength [30], form the building blocks for DNA nano-objects. The crossover connects this individual blocks (Figure 1.3A). The sequence specificity of complementary base-pairing enables chemical addressable self-assembly of artificial structure designs. In 2006, the method of DNA Origami was established by Paul Rothemund [1], when he created discrete structures containing up to half a million atoms by using a several thousand nucleotides long DNA strand (scaffold) that was folded into its tertiary structure by forming crossover with several shorter DNA strands. These shorter oligonucleotides (staples) are composed of sections with reverse-complementary sequences of different parts of the scaffold. On assembly, they "staple" the scaffold to form the outlined nanostructure. Figure 1.3B shows an exemplary design of the *Twisttower* object [32] with its scaffold folding topology (Figure 1.3C).
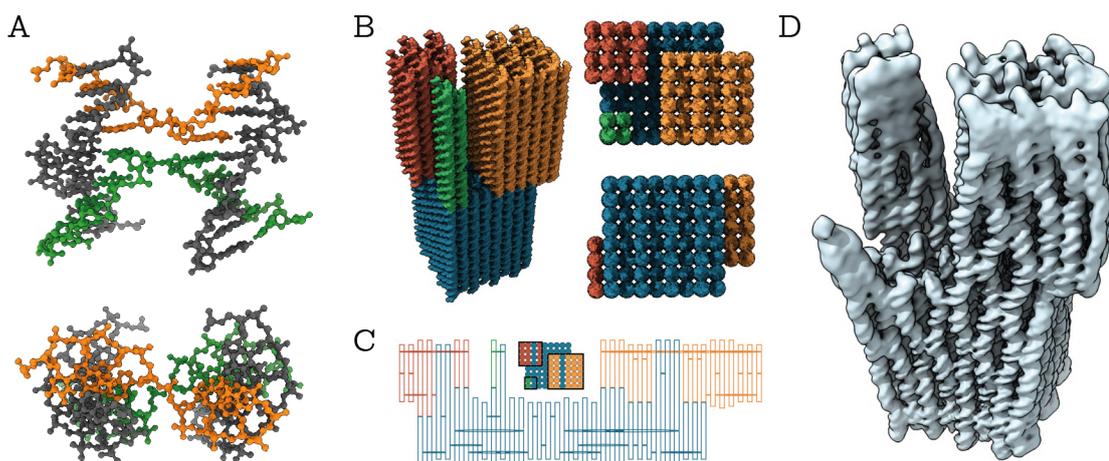
Figure 1.3 | *DNA Origami.* (A) *The helical crossover motif of the Holliday junction connects two double-stranded helices with strand-swap. The connecting oligonucleotides (blue) contain partially reverse-complementary sequences with each continuous oligonucleotide (orange) causing the formation of a helix duplex.* **(B)** *Combination of multiple junctions organized within a lattice enables the design of megadalton DNA complexes. The Twisttower object combines 4 domains of the varying cross-section in square-lattice.* **(C)** *To form a DNA origami nanostructure the scaffold strand is routed through the full structures. During self-assembly, it is folded into the programmed shape via the designed topology.* **(D)** *Experimental cryo-EM electron density data of the Twisttower Native Twist. Figure panels (B) to (D) adopted from F. Kohler [79].*

A more recent study [80] further solidified the suitability of the Holliday motif as the main junction motif in DNA nanotechnology. Their analysis of X-ray crystallography electron density data reveals only minor perturbations to duplex stability and solvent interaction at the junction compared to a regular duplex. The main conformational changes occurred in the torsional angles in the backbone, with the ξ-dihedral of the sugar-base connection (C1'-N1(Y)/N9(R)) switching from its syn- to anti-conformation and the β-dihedral of phosphate-sugar connection (O5'-C5') flipping around nearly completely. These two deviations in two successive nucleotides are sufficient to generate the extension and turn required to form the junction.

The use of three-dimensional lattices enables expanding the crossover network to form multilayered DNA origami structure [12, 26, 27, 81, 82] and triangulation-based wireframe DNA origami [6, 83, 84]. The techniques and protocols have since been optimized to allow more control over the self-assembly [85–88], increased stability against low salt and degradation [15, 17, 89], and mass production [90] of structures of ever-increasing size [2]. Development of custom scaffold sequences of multiple thousand nucleotides in length [91, 92] facilitates addressability of every single subdomain of the structure.

Application of DNA nanostructure range from macromolecular measuring instruments [5, 14, 37, 93] and preliminary stages of molecular machinery [10, 16, 18, 19] to biomedical utilization for drug delivery and vaccination [3, 13, 21, 22].

1.2   High Resolution Volumetric Cryogenic Electron Microscopy Data

Throughout this work, we evaluate and discuss volumetric electron density data obtained from cryogenic electron microscopy (cryo-EM). We use computational modeling (Chapter 2) to enrich the unstructured data with topological and positional information (Chapter 3), that allows structural evaluation and data mining (Chapter 4). We also utilize the experimental electron densities as the primary validation source for computational structure prediction (Chapter 5). Limitations and challenges associated with the generation of these datasets play an essential role in computational structure validation and prediction.

To add context to this data source, we give a brief overview of the fundamental principles of single-particle cryo-EM. State-of-the-art methodology for data processing of heterogeneous systems and challenges associated with imaging of megadalton DNA nanostructures are introduced. For a general introduction to electron microscopy, we recommend the book by J. Frank [94], which also serves as the main resource for this section. Publications associated with the *REgularized LIkelihood OptimizatioN* (RELION) software suite by S.H.W. Scheres and coworkers [95–100] represent the leading-edge of single-article analysis (SPA) research.

The overall dimensions of DNA origami nano-objects are typically around 30 nm (*Pointer Version 1* 32×26×44 nm) [33]. In contrast, important structural features are discernible only at the nucleotide level. To investigate the correct base-pair formation and local structure of the duplex, imaging at resolutions in the ångström-range are necessary, with the helical minor groove visible at around 12 Å or better [29]. Modern electron microscopy setups reach resolutions of up to 1.22 Å [100, 101]. While the theoretical resolution limit of the accelerated electron beam is defined by its small wavelength, experimental constraints of sample preparation and the lens/aperture setup of the microscope are the predominant limiting factors. Currently, the best-reported resolution for DNA origami nanostructures is 4.3 Å [32], with the biggest obstacle being the large heterogeneity of the imaged ensemble.

To avoid deformation of the structures during imaging and properly represent their native conformation in thermal equilibrium, they are vitrified in amorphous ice by freezing a thin layer of nano-objects in an aqueous solution in liquid ethane. The support grids that

contain these samples during the freezing are maintained at low temperatures for use in a cryogenic electron microscope. Besides properly preserving the native ensemble, cryogenic electron microscopy (cryo-EM) increases the resolution by direct imaging of the structures at a random orientation within the thin layer of ice.

In the microscope, a focused electron beam is scattered on the sample and collected at a detector plane. The scattering causes a phase shift $\phi(\mathbf{r})$ of the electron wave $\psi(\mathbf{r})$ which is dependent on the sample's Coulomb potential V.

$$\psi(\mathbf{r}) = \psi_0 \exp\left[i\phi(\mathbf{r})\right], \text{ with } \phi(\mathbf{r}) = \int V(x, y, z)dz \qquad (1.1)$$

At the electron detector, the intensity distribution of the electron wave $\psi(\mathbf{r})$ induces a signal from which the image of the sample can be derived. However, this signal and thereby the resolution of the final image are impaired as the electron wave is obstructed by apertures and distorted lens aberrations.

Each image generated by transmission electron microscopy represents a projection of the Coulomb potential of the sample perpendicular to the imaging plane. For the analysis of multilayer DNA origami nanostructures, the information provided by a single projection is insufficient. However, a combination of multiple projections of an object can be utilized to derive its three-dimensional structure via the central-slice theorem [102]. Besides the resolution of each image, the overall resolution of this 3D reconstruction is dependent on the number of projections and their angular distribution. In general, more images and a more uniform orientation distribution result in a higher resolution of the reconstruction.

When imaging bio-molecular materials like DNA constructs, the sample gets destroyed by the electron beam. Consequently, individual images come from different particles instead of one unique structure. In single-particle analysis with cryo-EM (SPA) numerous individual images are taken and collected to represent the native ensemble of the sample. The homogeneity of the sample is a major limitation for high resolution in SPA with strong deviations reducing the overall resolution. To counteract this effect sample preparation is highly optimized and multiple rounds of image classifications are performed on the projections. Only particles classified to represent similar configurations are used for the 3D reconstruction, while incomplete or deformed particles are excluded. Due to the one-pot self-assembly production of the DNA origami method [86], large numbers of individual

particles can be imaged in a single session. This approach, combined with the extensive particle classification allows mitigation of the sample inhomogeneity to a certain extent. Additionally, a good orientation distribution of the set of projections is achieved by plunge-freezing and cryogenic imaging.

Despite the advantages listed above, DNA origami objects have yet to reach resolutions close to SPA of protein [100] or X-ray crystallography of DNA crystals [103]. In contrast to these examples, scaffolded DNA origami nanostructures, composed of several thousand nucleotides, suffers two disadvantage. First, homogeneity is reduced by structural defects, as individual particles are compact and stable even without every single staple oligonucleotide present. Unbound nucleotides or synthesis defects like point mutations and terminal base deletions might further decrease sample uniformity. Second, DNA origami objects exhibit a large degree of internal motion due to their size and are often designed to occupy different conformations. In the case of hinge- or joint-like structural elements, the relative motion is a feature of the structure. However, the flexibility of the DNA duplex and the degrees of freedom of the Holliday junction allows for relative motion within the lattice. In Figure 1.3D the effect of domain motions within the same DNA origami design is most apparent in the reconstruction of the 2×2 domain which is hardly visible. Since regular SPA aligns the structure by the biggest overlap, the signal of the flexible 2×2 domain gets smeared most, significantly reducing the local resolution of this part of the nano-object and likewise the overall resolution. Although not completely disjoint sub-parts of the structure, the varying relative orientations of the four domains represent different overall conformations.

These different conformations of individual particles can be resolved with the multi-body refinement technique [99]. With this technique, the signal of individual domains is picked apart and reconstructed separately. The approach not only increases the overall resolution of the individual domain but is also capable of representing the relative motions of these domains. A similar technique called focused refinement [32] has been shown to mitigate the effect of lattice breathing motions of multilayer DNA origami and alignment artifacts.

1.3   Prediction & Validation of DNA Nanostructures

This section covers the principles of Molecular Dynamics simulation [104] and its utilization for structure prediction using all-atom as well as coarse-grained representation of DNA nanostructures. We also introduce the technique of MD flexible fitting, an MD-based approach for building pseudo-atomic models for experimental electron density data of macromolecules. For an extensive introduction the reader might refer to the books of

Rapaport [105], and *Frenkel & Smit* [106], which also serve as the main reference for this section.

1.3.1   Introduction Molecular Dynamics

A Molecular Dynamics simulation (MD) corresponds to the computation of the trajectory of a classical many-particle system using a discrete time step. The trajectory is generated by iteratively solving the equations of motion at every successive time step. By analyzing this time evolution a collection of thermodynamic properties and structural observables of the ensemble can be computed. Solving the equations of motion requires the position, velocity, and acceleration of each particle in the system. The forces responsible for their change are the result of a potential generated by various types of interaction between the particles in the system, coupling the motion of one particle to the motion of all others.

Within the Born-Oppenheimer approximation the Equations of Motion (EOM) for a classical system with N particles are:

$$\dot{\mathbf{r}}_i = \frac{\mathbf{p}_i}{m_i}$$
$$\dot{\mathbf{p}}_i = -\nabla_{\mathbf{r_i}} \mathcal{V} = \mathbf{F}_i, \tag{1.2}$$

with the force $\mathbf{F}_i$ exerted on particle i, the particle's position $\mathbf{r_i}$ and its momentum $\mathbf{p_i}$. The system of $6N$ first order differential equations is coupled by the gradient of the potential energy $\mathcal{V}$. In an MD simulation, these equations are solved numerically in discrete time intervals Δt, computing the trajectory of the system. To propagate particle positions and velocities the forces acting on each particle are determined at each time step by evaluating their potential energy. In the case of an ergodic system, this time-evolution will adequately sample all its accessible microstates. The fundamental postulate of statistical mechanics states the equal probability of any microstate of an isolated system. Following these propositions leads to the ergodic hypothesis: The temporal average $\bar{A}_t$ of a trajectory equals the ensemble average $\langle A \rangle$ of the macrostate for any observable $A(\mathbf{r^N}, \mathbf{p^N})$.

$$\lim_{t \to \infty} \frac{1}{t} \int_0^t d\tau A(\mathbf{r^N}(\tau), \mathbf{p^N}(\tau)) = \int d\mathbf{r^N} d\mathbf{p^N} f(\mathbf{r^N}, \mathbf{p^N}) A(\mathbf{r^N}, \mathbf{p^N})$$
$$\bar{A}_t = \langle A \rangle, \tag{1.3}$$

with probability density $f(\mathbf{r^N}, \mathbf{p^N})$ of the microstates $(\mathbf{r^N}, \mathbf{p^N})$.

However, numerically computation of the EOM is generally Ljapunov instable, meaning an exponentially increasing error propagation [107]. Consequently, numerically computed trajectories do not correspond to the real trajectories of a particle due to the finite accuracy of computer hardware and integration algorithms. The longer a trajectory becomes, the larger the deviation between computed and real microstate gets. Fortunately, MD simulation focus on the accurate sampling of a macrostate rather than a singular trajectory [108]. It is possible to show, that for the group of symplectic algorithms the probability density function of such a trajectory only deviates by $\mathcal{O}(\Delta t^n)$, where n is the order if the integration algorithm. The use of such an algorithm hence guarantees accurate sampling of the ensemble.

With the calculation of forces being the computational most expensive at $\mathcal{O}(N^2)$ for a system of N interaction sites, algorithms that require fewer evaluations of the interaction potentials are generally preferred. The 2nd-order velocity Verlet [109], a self-starting and memory-efficient variant of Størmer-Verlet [110] algorithm.

$$
\begin{aligned}
\mathbf{r}(t + \Delta t) &= \mathbf{r}(t) + \mathbf{v}(t)\Delta t + \frac{\mathbf{F}(t)}{2m}\Delta t^2 + \mathcal{O}(\Delta t^3) \\
\mathbf{v}(t + \Delta t) &= \mathbf{v}(t) + \frac{\mathbf{F}(t) + \mathbf{F}(t + \Delta t)}{2m}\Delta t + \mathcal{O}(\Delta t^2)
\end{aligned}
\tag{1.4}
$$

The velocity Verlet algorithm is symplectic and can be derived by combining the regular and time-reversed Taylor expansion of the position. Despite its accuracy of only 2nd-order, it is one of the most popular integration algorithms for MD since it allows proper sampling of the equilibrium ensemble of bigger systems.

The forces exerted on interaction centers during an MD simulated are determined by interaction potentials. The collection of different interaction potentials present in a system is called its force field. Similar to the choice of integration algorithm, the development of these force fields is focused on maximizing computational efficiency without compromising the accuracy of the modeled physical interaction. Their functional form is often derived from theoretical considerations or modeled from experimental data or Quantum Chemistry simulations. In general, interaction potentials with exponential terms, square roots, and trigonometric functions are avoided as modern hardware is highly optimized for addition and multiplication. One example is the regular use of harmonic oscillators to model covalent bonds. Their evaluation requires only one addition and two multiplications while representing the bond reasonably well and only requiring two parameters. Another common example is the Lennard-Jones 12–6 intermolecular pair

potential [111, 112]. It models short range Pauli repulsion via a $-\frac{1}{r^{12}}$ term and long-ranged attraction of dispersion forces by a $\frac{1}{r^6}$ term. Notably, only the attractive part is based on theoretical considerations of London dispersion forces, while the repulsive part has been chosen for computational efficiency. For the particular case of DNA several iterations of force fields have been published [53–56, 59, 113, 114], with CHARMM:36 [75] and AMBER:OL15 [115] most commonly used today. The choice of a particular force field might depend on both the specific system studied along with the simulation framework used.

The general scheme for any simulation framework can be separated into several main components, represented in Algorithm 1. First, the system has to be initialized, including allocation of memory and parsing of initial conformation, simulation parameters, and force field specifications. Simulation parameters include among others the time step Δt and the number of iteration steps M to be executed. For each of the M iterations, forces for the current conformation are computed according to the force field. Next, the trajectory comprising position and velocity data of the system are propagated to $t + \Delta t$ by solving the EOM using a finite difference integration algorithm. Lastly, thermodynamic averages, like temperature and pressure, have to be updated according to the conformation of the system at the subsequent time step. Simulations presented in this work are performed in the canonical ensemble, with the temperature being controlled by coupling the system by the first-order response of a thermostat.

Algorithm 1: Basic MD structure

Input: simulation parameters: Δt, *M, N*

Initialization();
for *M steps* **do**
    forces = **compute_forces()**;
    trajectory += **integrate_EOM**(Δt, *forces, integrator*);

    [] *averages* = update_averages();
end
return *trajectory, averages*

To further increase the performance of any MD simulation several tricks are employed throughout the main loop. For instance, interactions are usually only evaluated for interaction pairs that are in reasonable proximity. Appropriate cutoff distances are strongly dependent on the functional form of the interaction potential and require either a negligible contribution beyond the cutoff or an analytical long-range correction term. The utilization of neighbor-lists [110] that track potential interaction pairs for several time steps combined with these interaction cutoffs helps reduce the computational cost significantly.

Finally, initial configurations are often derived from crystal structures or equilibrium conformations of a similar macro state. MD simulations therefore usually start in a thermally excited state and have to be relaxed first. Several computational schemes have been developed to efficiently reach thermodynamic equilibrium, the DNA origami-specific example *mrDNA* [116] is described in more detail in the next subsection. Trajectories and thermodynamic averages computed throughout this process of relaxation are usually discarded as they do not sample microstates of the correct ensemble.

1.3.2 Computational Structure Prediction

During the design process of multilayer DNA origami nanostructures, the topology of a structure is generated based on an idealized crystal lattice toy model [86]. To ensure the stability of a nano-object, helices have to be sufficiently connected to helical neighbors via a strand-swap crossover. The choice of a specific subset of connections from the pool of possible connections between a helix pair determines whether all subdomains of the object are adequately stabilized. Apart from these topological requirements for the overall structure of a DNA origami design, the abstraction level of the origami lattice causes deviations from the true shape of experimentally validated DNA complexes. The model for double-stranded B-DNA used in these lattices requires the discretization of double-helical segments into simplified building blocks. Generally, 21 base-pair building blocks comprise two full helical turns in the lattice, with discontinuities in the backbone due to nick sites or Holliday junctions treated structurally equivalent to fully double-stranded DNA segments. Not considered in the model are slight dissimilarities in helical pitch along with the position of the backbone as influenced by helical grooving. The placement of crossover motifs forms the core element of the topological connectivity of neighboring helices, thereby introducing internal stresses and torques. Moreover, short staple-segments might not be stable enough under experimental conditions and never form the duplex intended by the designed topology. The staple-segment distribution and staple-breaking rules [85] also have a strong influence on self-assembly. These effects can lead to both local and global deformations of the nanostructure, requiring an iterative refinement workflow for the design of DNA origami nanostructures.

Throughout the iterative development process, designers use experimental and computational measures to validate or detect issues with their design and update the topology accordingly. While experimental methods like gel-electrophoresis and electron microscopy are indispensable as a final means of validation, they are also time-consuming and expensive. Computational structure prediction is fast and considerably less expensive. It can provide

insights into several of the issues listed above throughout the design process. Additionally, the visualization provided by the prediction itself might help guide the design process of the macromolecule.

The trajectory of an MD simulation can be used to compute a variety of average quantities that can be a great resource in assessing a design. While the trajectory itself might not represent the real time-evolution of a single particle, it generates an ensemble of viable configurations for the specific model. This set of data points can be used to compute the average structure and quantify the root-mean-square fluctuation of individual parts around this average. Analysis of this data enables quantifying the overall dimensions and shape and provides a visualization of the phase space available to the structure. If the structure can populate several distinct conformations or exhibits strong domain motions, the trajectory can supplement additional context to the DNA origami design process.

Coarse Grained DNA Models

With up to $5 \cdot 10^6$ atoms for typical multilayer DNA origami, all-atom MD simulations remain at the limit of modern hardware [31]. This is especially true for slow domain motions or solvent interaction, due to the required timescale or the number of solvent molecules, respectively. However, most nanostructure designs do not require atom-level resolution to validate design goals. By reducing the number of interaction sites the computational cost of a simulation can be reduced significantly. This process, called coarse-graining, defines interaction centers that combine the properties of groups of atoms, with specifically tailored interaction potentials to phenomenologically represent the system. Importantly, the coarseness of the model is balanced with its predictive power, with a finer model able to simulate more intricate effects.

Here, we give a brief overview of the most popular coarse-grained structure prediction models for DNA origami nanostructures, listed from finest to coarsest.

Elastic Network-Guided Molecular Dynamics

The coarse-graining procedure does not necessarily need to be carried out on the molecule of interest itself. For Elastic Network-Guided Molecular Dynamics (*enrgMD*) [31] it is the atoms of the solvent and their effect on the DNA origami that is reduced to a phenomenological representation. While maintaining an all-atom representation of the DNA strands, a network of elastic bonds replaces the solvent 1.4A. The purpose and benefits of the elastic network are multifold. Firstly, this implicit treatment of the solvent reduces the computational cost by several orders of magnitude, by removing the necessity
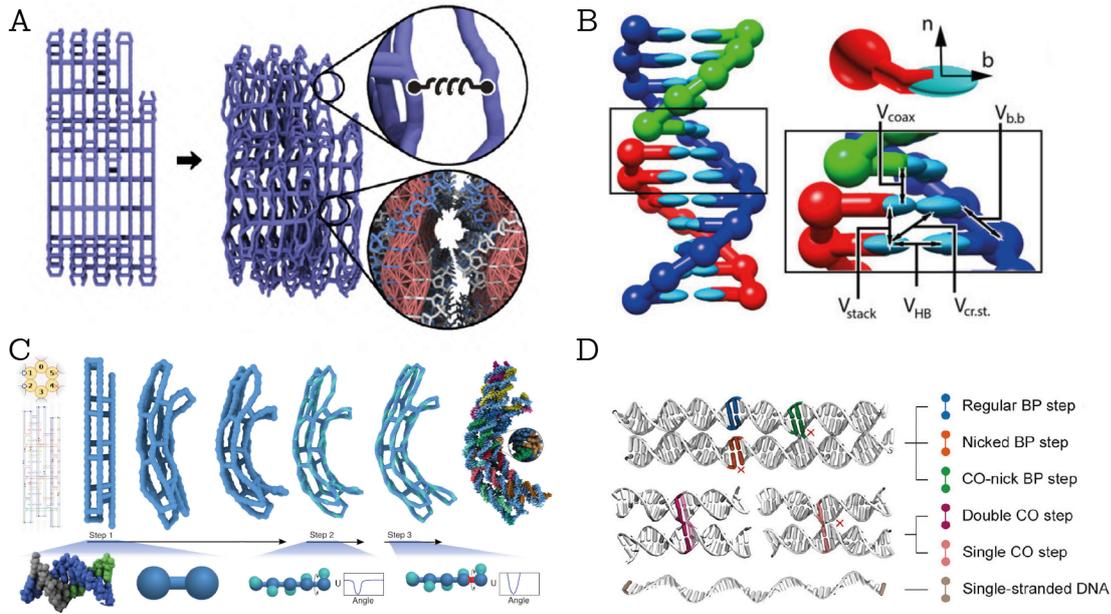
Figure 1.4 | *Coarse-Graining of DNA origami nanostructure simulation: (A)* For *enrgMD, the solvent molecules are replaced with a network of elastic bonds, while the molecules of the DNA nano-object retain their atomic level representation. Adopted from [31]. (B) In the oxDNA model a nucleotide is represented with one interaction site located at the nucleobase and another one on the backbone. Adopted from [117]. (C) Gradual increase of granularity by increasing the number of interaction sites of a mrDNA relaxation. Adopted from [116]. (D) Structural motif classification of SNUPI. Adopted from [39].*

to explicitly solve the EOM of the water and salt molecules as well as their contribution to the potential energy of the DNA strands. Secondly, harmonic bonds inside the DNA duplex stabilize its structure and prevent the dissociation of base-pairs by mimicking the effect of the solvent. Thirdly, elastic interactions between helices that are adjacent in the origami lattice emulate long-range electrostatics, as the repulsion of the DNA backbones is not entirely shielded by the solvent. Since these long-range effects are covered by the network, a smaller cutoff can be used for the computation of the interaction term. Finally, the increased stability of the tertiary structure of the DNA nano-object facilitates the choice of a larger time-step for the integration operator. These faster dynamics enable better relaxation and improved sampling of the equilibrium ensemble.

To produce accurate predictions of these effects the elastic model was parametrized using a single explicit solvent all-atom simulation of the *Pointer Version 1* object [33] and validates against the pseudo-atomic model of its cryo-EM data. The parameter-set of the network for the *Pointer Version 1* object generalizes well to simulations of other structures. This enables fast structure prediction without compromising accuracy, with a cumulative speedup of four orders of magnitude [31] compared to explicit solvent MD.

oxDNA Model

In the *oxDNA* model [118] each nucleotide of a DNA strand is represented by two interaction centers. The inner site is located at the position of the nucleobase with interaction terms for base-pairing and base-stacking. The outer site models the backbone with connectivity within the strand and electrostatic repulsion of the negatively charged phosphate group. In its updated version [76], the *oxDNA* potential is composed of a total of eight interaction terms, with three targeting next-bonded neighbors and five targeting non-bonded pairs:

$$
\begin{aligned}
V = &\sum_{n.n.} \left(V_{bb} + V_{stack} + V'_{exc} \right) \\
&+ \sum_{pairs} \left(V_{pair} + V_{cross-stack} + V_{coax-stack} + V_{DH} + V_{exc} \right)
\end{aligned}
\tag{1.5}
$$

The backbone connection of bonded neighbors V_{bb} is modeled using a finitely extensible nonlinear elastic spring. Hydrogen bonding of complementary nucleobases, V_{pair} orbital $\pi - \pi$ stacking of adjacent bases within a strand V_{stack}, coaxial stacking of non-neighboring bases $V_{coax-stack}$ and cross-stacking between base-paired neighbors $V_{cross-stack}$ enable prediction of the structural features of a DNA duplex. Helical grooving is achieved by tuning the angle between the hydrogen-bonding and stacking site with the slightly offset backbone site. Volume-exclusion terms for neighbors V'_{exc} and non-bonded pairs V_{exc} prohibit steric clashes. The electrostatic interaction V_{DH} of the backbone is based on the Debye-Hückel model. This electrostatic potential adds temperature- T and salt-dependence I to the coulomb potential using the Debye screening length λ_D:

$$
\lambda_D(I,T) = \sqrt{\frac{\epsilon_0 \epsilon_r k_B T}{2 N_A e^2 I}},
\tag{1.6}
$$

with the vacuum permittivity ϵ_0, the relative permittivity of water $\epsilon_r = 80$, Avogadro's number N_A, Boltzmann's constant k_B, and elementary charge e.

Accurate modeling of the backbone repulsion is particularly important for DNA origami nanotechnology, where solutions of high ionic strength are required to allow compact packing of multiple helices within the lattice.

As discussed in the previous Section 1.3, interaction potentials are chosen for their functional form to reproduce an experimentally observed or theoretically predicted effect. Exemplarily, the stacking potential V_{stack} of adjacent bases, which is essential for duplex-formation, is implemented using a Morse potential of the form:

$$V_{stack}(\mathbf{r}) = \epsilon \cdot \left(1 - e^{-a \cdot (r-\sigma)}\right)^2, \tag{1.7}$$

with $\mathbf{r}$ the distance. The three tunable parameters a, ϵ, and σ affect width, height and position of the potential well. Notably, the Morse potential contains an exponential term but is considered more accurate for diatomic bond modeling than the Lennard-Jones potential. A potential of the same form is used for the sequence-dependent hydrogen-bonding V_{pair}. In contrast to a harmonic bond, the Morse potential allows for bond breaking, enabling bases to hybridize and stack dynamically.

The *oxDNA* model reduced the number of interaction sites by a factor of roughly 15 for DNA and treats salt-dependent electrostatic screening of the solvent implicitly. Compared to all-atom MD simulations this drastically reduces the computational cost, allowing for structure predictions of DNA origami with up to 10,000 base-pairs within a couple of hours.

The code for *oxDNA* is implemented in C++ [38, 119] and parallelized for GPU [120]. Several complementary software and web services for simulation (*tacoxDNA* [121]) and analysis (*oxView* [122]) of DNA nanostructures are available. Initial models for simulation of DNA origami can be generated directly from the caDNAno strand diagram and the strand sequences. Adaptation of the model has led to the development of the compatible models oxRNA [123, 124] for RNA and an anisotropic network model for equilibrium fluctuations of protein components [122].

Multi-Resolution DNA Simulation Framework

Besides the simulation of the dynamics of a system and the proper sampling of the accessible phase-space, significant resources need to be expended on the preparation of a valid starting configuration in thermal equilibrium. Relaxation of structures that can only be prepared far away from valid configuration is particularly computationally intensive and might require several steps of manual intervention. For instance, the prediction of a DNA origami object with a corner motif (two lattice-based subdomains fused at an angle) requires pre-alignment of the two subdomains and careful relaxation of bonds in the corner region. This holds for both *enrgMD* and *oxDNA*, with *oxView* providing specialized alignment tools. A similar situation occurs for designs that deviate strongly from the lattice, such as highly curved structures. These situations cannot be resolved by pre-alignment and require long relaxation runs.

However, in the early stages of relaxation towards a viable solution structure hardly any detail is necessary. With the Multi-Resolution DNA Simulation Framework (*mrDNA*) [116]

this instance is exploited by computing several coarse lattice-based DNA origami models of varying granularity. Each model can be converted into the adjacent layers of abstraction. This facilitates, among other protocols, the creation of a fast relaxation pipeline.

The *mrDNA* framework is based on the Atomic Resolution Brownian Dynamics [125] bead model (ARBD). At its coarsest stage, multiple base-pairs are represented by a single bead with limited interaction potentials. As the simulation progresses the granularity of the model is increased by introducing more beads, while also increasing the complexity of their interactions. For relaxation to thermal equilibrium, the model becomes less coarse with each successive stage (Figure 1.4C).

The ARBD bead model is optimized for GPU and can generate both *oxDNA* and *enrgMD* simulation files. Due to the coarseness of the initial stages and the flexibility of the framework, nearly all DNA origami designs can be efficiently and accurately relaxed, even without requiring manual intervention. This approach drastically reduces the computational cost and enables the relaxation of designs that would not relax reasonably well in *oxDNA* or *enrgMD* alone. Additionally, the framework is developed open-source in Python making it highly customizable. It also enables the study of the dynamics of DNA nanostructures in the presence of electric fields and other conditions.

Finite Elements Structure Prediction of DNA Origami

An alternative to MD simulation is the computational prediction by a structural finite element (FE) model tailored to lattice-based DNA origami nanostructures. While potentially not as accurate as the previously introduced techniques, FE allows the incorporation of sequence-dependent properties and geometries of structural motifs typical for DNA origami nano-objects into a comparatively coarse model. These models are therefore valuable tools for fast prototyping, iterative design, and structure prediction.

For *SNUPI* (Structured NUcleic acids Programming Interface) [39] and its predecessor CanDo [26, 126, 127], the DNA object is divided into base-pair steps which are differentiated by the type of motif they represent. In the case of *SNUPI* these are regular and nicked base-pair steps, various types of crossover motifs, and single-stranded DNA (Figure 1.4D). Each motif is further discriminated by sequence, constituting the complete set of available elements. The parametrization of the relative geometries of these elements and their structural properties are derived from MD simulations and compared to experimental data. By implementing a partition & relocation framework [128], *SNUPI* is capable of predicting the shape of multidomain DNA origami and DNA Wireframe structures. The structure

prediction with the FE approach can be computed on the timescale of a few minutes on standard hardware.

1.3.3   Pseudo-Atomic Model Building

Electron density distributions of macro-molecules measured by cryo-EM or X-ray crystallography are volumetric data sets, that do not explicitly contain any categorization or annotation. Building a Pseudo-Atomic Model (PAM) is a means to enrich this data with topological and positional information, along with insights about its constituting molecules. The resulting structures can be used for exploratory data analysis [129] and the investigation of their physical properties [34, 35, 130].

Historically, quantitative rigid-body docking [131–137] of models into low- to mid-resolution electron density data was used to determine solution structures. However, the inherent flexibility of larger molecular complexes required the development of a more proficient methodology [138]. By division into rigid units, flexible rearrangement [36, 139], conformational sampling [140] and deformations [141, 142] of the model enables better overall PAMs. The main drawback of these approaches is the introduction of ambiguity by the choice of subunits and the potential creation of nonphysical conformations. These can occur by the neglecting internal flexibility of the individual domains and coordinated domain motions of the system.

A possible solution for these challenges is the inclusion of MD simulations within the fitting process [130, 143]. For these methods, all atoms of the model are steered by a biasing potential to refine and improve the alignment of the PAM and the electron density data. Since the properties and dynamics of the full structure are described by the MD simulation, nonphysical results can be avoided. The designation of rigid elements is no longer required and conformational changes are governed by the interaction potentials of the MD force field. Full flexibility of the macromolecule comes at the expense of increased computational cost associated with the MD simulation but is indispensable, especially at higher resolutions.

A significant benefit of a fitting approach based on a biased MD simulation is the independence of its success rate from the size of the system. The introduction of the steering potential [144], does not significantly increase the computational complexity of the simulation and the optimization of the structure is performed locally for each atom. Contrary, the global optimization criterion of rigid body-docking and stochastic optimization increases with the number of degrees of freedom in the system[130].

This property of MD-based model building is especially important for the field of DNA nanotechnology as structures are often on the megadalton scale. Additionally, high-resolution cryo-EM of DNA origami nano-objects has revealed strong internal motion and deformation within the origami lattice. The internal degrees of freedom, without the presence of distinctly identifiable subdomains, make conventional fitting techniques unsuitable.

Molecular Dynamics Flexible Fitting

Molecular Dynamics Flexible Fitting (MDff) was originally developed by the group of Klaus Schulten in 2008 [130] and remains a key technique for atomic model building. Many advanced protocols [138, 145–149] are based on its fundamental principles. At its core, MDff is a steered MD simulation [144, 150] with a grid-based potential [151] that is based on experimental cryo-EM data (compare Figure 1.5A). Since electron density data are stored as a volumetric data set, no experimental information is lost during the computation of the grid.
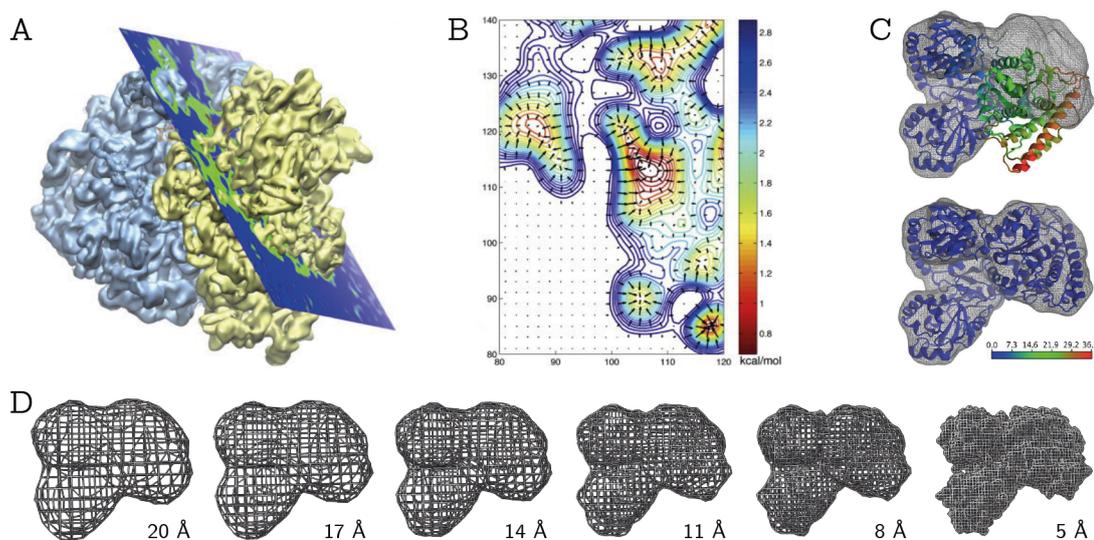
Figure 1.5 | MD flexible fitting: (A) *Cryo-EM data of the ribosome at an overall resolution of 6.7 Å [152] depicted via an iso-surface. The plane represents an arbitrary 2D slice of the volumetric dataset.* **(B)** *Representation of the grid-based potential generated from the slice depicted in A. Colors depict the strength of the potential, with intensities colored by a rainbow palette. Small arrows indicate the direction and strength of the resulting forces at each grid point.* **(C)** *Initial configuration (top) and final result (bottom) of MD flexible fitting of acetyl-CoA synthase (PDB 1OAO) of the ribosome data presented in (A).* **(D)** *Cascade of six cryo-EM maps of the acetyl-CoA synthase with increasing resolution. Overall resolution ranges from 20 Å (left) to 5 Å (right). Maps were generated from atomic coordinates of (PDB 1OAO) using ChimeraX. (A-B) adapted from [145], (C) from [130].*

Acquired by cryo-EM, the volumetric data represents the electron distribution of the macromolecule and solvent. While the contribution of the bulk solvent is smeared by SPA, high-intensity regions of the volumetric data represent the Coulomb potential $\Phi(\mathbf{r})$ of the macromolecule. Its contribution to the potential energy $U_{(EM)}$ can be defined as:

$$
U_{(EM)} = \sum_j^N w_j \begin{cases} \zeta \left[1 - \dfrac{\Phi(\mathbf{r}_j) - \Phi_{(thr)}}{\Phi_{(max)} - \Phi_{(thr)}} \right] & \text{if } \Phi(\mathbf{r}_j) \leq \Phi_{(thr)}, \\ \zeta & \text{else.} \end{cases} \tag{1.8}
$$

The total potential collects contributions of all N atomic coordinates $\mathbf{r}_j$ of the system, with the weight w_j accounting for the type of atom — usually via the atomic mass. Signal not corresponding to the target molecule are neglected using the threshold value $\Phi_{(thr)}$ and subsequently normalized using the maximum value of the electron density data. The difference-term reveals the nature of the potential as attractive towards high-intensity regions, as shown in Figure 1.5B. Consequently, atoms will drift such that they match the cryo-EM signal as much as permitted by the MD force field (Figure 1.5C).

The scaling factor ($\zeta > 0$) tunes the strength of the resulting force globally, while w_j modulates it on a per-atom bases. The choice of the proper scaling factor requires a balance of fitting success with a stable trajectory and prevention of overfitting. Strong accelerations of single atoms caused by large values of ζ might cause a crash of the MD simulation and potentially force a nonphysical conformation. The grid-force potential must not exceed values that distorted the macromolecule beyond the boundaries provided by the MD force field.

An additional measure against overfitting [153] is the use of artificial restraints that preserve the stereochemical quality of the system. Typically, dihedral angle constraints and harmonic restraints for hydrogen-bonded atoms are used early in the fitting procedure. Once approaching the target conformation, the restraints can be released for domains with sufficiently high electron density resolution. Other techniques like simulated annealing are also employed to overcome the energy barriers of local minima. Multi-step protocols, that combine these different methods are commonly developed based on the characteristics of the targeted macromolecules.

Generally, the MDff technique is extremely versatile and adaptable. In addition to the compilation of different sets of restraints and simulation parameters, flexible fitting can be performed for only a subset of atoms in the system. Hydrogen atoms for instance are generally omitted from interaction with the grid-based potential as their contribution to

the solution structure is negligible. This is possible because the optimization of MDff is performed locally, due to the sole dependence of the grid-force term on the gradient at the particular grid point. The fitting procedure can therefore be applied for any component of the molecule while maintaining the complete topological context of the complete system.

Quality Assessment of Pseudo-Atomic Models

Most use cases of PAM building and flexible fitting revolve around the model building for previously characterized structures or conformations. For these structures, all components of the molecule and their topology are known, facilitating quantitative comparison of previously constructed models. However, assessment of the quality of the solution conformation is impeded by the data type of the volumetric electron density and its inherently noisy signal. The most commonly reported metric is the so-called global cross-correlation coefficient of model and cryo-EM data. Here, an artificial electron density map is simulated based on the PAM [154] and computed at the same voxel grid and overall resolution of the experimental data. The final metric is the Pearson's correlation coefficient over all voxels of the pair of experimental and artificial datasets. The term *global* refers to the complete set of voxels being used for the computation.

Since the global coefficient depends on all voxels of the experimental data, the background signal of the bulk solvent and noise also contribute to the overall value. Yet, the size of the volume box is specified by the electron microscopist, with larger volumes arbitrarily inflating the correlation metric. To mitigate this effect and more accurately assess the qualify of the fit, different measures for local cross-correlation coefficient [155] have been proposed. In the case of masked cross-correlation [156], only voxels that are within a certain radius of the atomic centers are included in the computation. Other alternative measures are the computation of a minimal rectangular box, the inclusion of voxels inside the molecular envelope defined by a threshold value of the electron density, or only regions with the highest overall intensities. While the first example only partially resolves the problem of an arbitrary signal contribution of empty box regions, the latter is particularly well suited for macromolecules with a potentially unclear composition. In all cases the region of interest is derived from the exterminate data, wherefore electronic signals that are not matched by the electronic model are particularly penalized [157]. For masked cross-correlation, the mask is computed from the PAM causing a decrease in the case of the model containing atoms that are not reflected in the experimental data. This characterization makes masking highly suitable for DNA origami nanostructures where all nucleotides of the structure are fully characterized in advance. Correlation metrics can also be included directly in the fitting process by constantly updating the MDff potential according to the correlation of the current conformation [149]. However, the computational

cost associated with the simulation of the artificial electron density renders these techniques impractical for large structures.

In addition to correlation metrics, the stereochemical quality of the macromolecule can be assessed to evaluate the quality of a fitted pseudo-atomic model. The tools provided by the web server MolProbity [34] and the Protein Data Bank (PDB) [158] are considered community standards, but the geometric information provided by the model also enables the generation of custom quality metrics.

Cascaded Fitting Procedures

A significant disadvantage of MDff is its reliance on an exclusively attractive grid-force potential. This is particularly problematic if the initial model deviates strongly from the target structure. As the grid-based potential is indiscriminate of the molecular identity and its context within the macromolecule, subdomains can get stuck in the wrong region of electron density during the initial alignment phase. The local minimum in the energy landscape of the cryo-EM data frustrates the flexible fitting. Also, strong accelerations caused by sharp intensity peaks can make the simulation unstable during the first few stages of the fitting protocol.

Not only initial alignment but also subdomain realignment can be drastically improved if low-pass filtered electron density data are used instead of data at the original resolution. Global alignment of subdomains predominantly relies on the long-range information present in the experimental dataset. Temporary removal of the high-resolution information during the early stages of the fitting process facilitates the avoidance of local minima. The overall quality of a PAM can be increased by using a cascaded relaxation protocol (cMDff) [147]. Here, MDff starts with a low-resolution map, which is gradually replaced with a map of higher resolution once the alignment advances (Figure 1.5D). Using this approach, alignment is staggered from global to a more local and model arrangement.

In general, cMDff belongs to a group of methods [130, 147, 148, 159] that decouple various aspects of an MDff simulation — like cryo-EM resolution, simulation temperature, or the grid-scaling factor — from the fitting target. The underlying idea behind these methods is harnessing the diverse benefit of multiple distinct stages during the fitting protocol. The simulated annealing step of the original protocol published with the MDff technique [130] is an early example of this approach, stating its capability to avoid local minima.

Chapter 2

Pseudo-Atomic Model Building

Remark: The work presented in this chapter was published in Feigl, Kube, Kohler, et al. Revealing the structures of megadalton-scale DNA complexes with nucleotide resolution (2020) [32]. Kube, M. and Kohler, F. performed cryo-EM measurements and data processing of the volumetric data presented here. Electron microscopy data not published there will be marked accordingly.

Pseudo-atomic model construction enables structuring and quantifying electron density data of macro-molecules [34, 35, 160]. In recent years, the availability of high-resolution X-ray crystallography imaging techniques [161] and cryogenic electron microscopy (cryo-EM) has caused a near exponential increase in imaged macro-molecules [162]. Simultaneously, computational methods for building pseudo-atomic solution structures have been developed [36, 130, 141, 157, 163]. However, in contrast to atomic modeling of proteins, few of these methods are suitable for modeling DNA nanostructures due to the size, of these objects and the pseudo-symmetry of the origami-lattice.

One method that is scalable for large structures and adaptable for all macromolecules is Molecular Dynamics based flexible fitting (MDff) [130]. The universal methodology of MDff has seen constant improvements and modifications [130, 138, 145–147] in recent years. The currently published cryo-EM density maps of DNA origami nanostructures [32] reach resolutions of up to $5\,\text{Å}$ using cryo-EM. This nucleotide-level resolution is high enough to render the association of the volumetric data with atomic-level models possible [159].

Here, we present a methodology specifically tailored to construct pseudo-atomic models (PAM) for DNA origami nanostructures using a semi-automatic protocol based on the technique of MDff. Most nanostructures assemble from a total of around 16 000 nucleotides [86]. Consequently, the number of interaction sites present in the structures is approximately

half a million. As computational cost scales with the size of the model, a combination of multiple steps combining different methods and strategies is necessary to create a fast and accurate protocol.

In this chapter, we first describe the developed method in detail, discussing the necessity and benefits of each part. Second, we will present a library of PAMs of DNA origami nanostructures. Finally, we introduce our DNA origami model building framework *dnaFit*, a Python-based pseudo-atomic model building and solution structure analysis framework.

2.1   Shrink-Wrap Fitting

At its core, an electron density map presents a three-dimensional volumetric dataset and hence inherits challenges associated with this data type. Notably, conventional computer monitor hardware still requires the representation of three-dimensional data as a two-dimensional image. A suitable representation is dependent on both the type of the raw data and the analysis goal. Manipulation of volumetric data is computationally expensive due to its cubic scaling. As the dimensionality of the represented data increases, the association of symbolic and spatial aspects of its information enables complexity reduction [164]. One efficient way to facilitate image interpretation is through enrichment of the data with additional context. Linking the volumetric data to preexisting information helps to annotate, segment, and structure its content.

For volumetric electron density maps, pseudo-atomic models can provide this link. The main obstacle to generating these models for common DNA nanostructures is their size. The protocol *Shrink-Wrap Fitting* is a combination of several existing methods and strategies. This way, we can utilize the benefits of each method and couple their strengths with particular features of DNA origami designs. Consequently, our methodology is tailored to the field of lattice-based DNA origami nanostructures.

The term *Shrink-Wrap Fitting* is a figurative synonym for the method's full descriptive name *Multi-Resolution Elastic Network-Guided Cascaded Molecular Dynamics Flexible Fitting*. While the full name is cumbersome to write, it points at the nature of this protocol.

2.1.1   Protocol

First, an initial pseudo-atomic model is generated using the multi-resolution framework *mrDNA* [116]. With the caDNAno strand diagram and the nucleotide sequence as input, this method creates a structure prediction using increasing levels of detail (also see

Section 1.3). This technique allows the creation of initial models for a broad spectrum of lattice-based DNA origami, including structures consisting of multiple subdomains. After the final step prediction step, *mrDNA* outputs an atomic-detail model supplemented with an elastic network produced by *enrgMD* [31]. The elastic network helps to increase simulation speed while retaining the structural stability of the nano-object. It also allows simulating without a solvent, further reducing the computational cost.

Before starting the fitting procedure, the initial model has to be roughly aligned with the cryo-EM map. This step is required to increase the computational efficiency and ensures that the globally optimal solution structure can be determined. Currently, manual pre-aligned is the only remaining non-automated step in the protocol. As simulations are performed in NAMD2 [165] we recommend VMD [166] for this step.

During the following steps, the initial model is systematically adapted to the electron density data obtained from cryo-EM. The atomic coordinates are corrected until they fit the experimental data. Performing accurate structure prediction before the actual fitting procedure helps decrease simulation time, as fewer changes have to be made during the adaptation. While not the fastest, *mrDNA* currently presents the most accurate structure prediction of DNA origami nanostructures [116]. Its seamless integration with *enrgMD* reduces computational overhead significantly. At this point, we like to highlight the symbiotic connection between structure prediction and model building. On the one hand, pseudo-atomic models derived from experimental data serve as the basis for parametrization of structure prediction model [31, 39, 118, 167]. The availability of more and better models will thereby inevitably improve structure prediction. On the other hand, the fitting of PAMs into experimental data can be facilitated and enhanced by using these improved predictions.

The fitting procedure is performed using molecular dynamics flexible fitting (MDff) [130]. The cryo-EM data are converted into a grid-based potential [151] which acts in form of an attractive steering force on all interaction sites of the electron-density map. Voxels with a stronger signal, which correlate to a higher electron density, cause a stronger attraction. Consequently, atoms will be drawn towards positions of high intensity in the experimental data. Simultaneously, both the *enrgMD* elastic network and the force field of the MD simulation restrict the motion of atoms to retain the topology.

The pseudo-periodic lattice structure of the volumetric data creates local minima that increase the risk of helices converging to the wrong position. Local minima are especially problematic if the initial docking is not optimal or the predicted dimensions of the PAM do not match the experimental data. To mitigate these issues and increase the chance

of fitting the global minimum configuration, we use a cascaded relaxation protocol [147]. Here, the electron density data that is used to generate the grid-based potential is initially low-pass-filtered to produce a low-resolution map with values of 25 Å or above. At this resolution, individual helices can no longer be discerned, allowing proper alignment of the whole structure. The low pass filtering is reduced stepwise until the original resolution is reached again and the correlation between the cryo-EM data and the model increases. If the initial docking is particularly poor, it can be beneficial to stop the cascade after the first few stages and restart the cascade of low-pass-filtered maps. Additionally, a short simulated annealing [106] phase can be included halfway through the cascade. The increased thermal fluctuations caused by the heating can help subdomains escape local minima. With these techniques, small misalignments that stem from deviations between prediction and measured data or insufficient docking can be resolved.

The elastic network provided by *enrgMD* has two different beneficial effects on the fitting procedure. First and foremost, it allows a significant speedup by enabling solvent-free simulation and using a longer time step. Yet, it also ensures structural stability of the DNA nanostructure throughout the simulation. In the early stages of the cascade, the model might deform significantly, as helices need to bend or move several nanometers to reach their target location. Since strong forces might act on the individual atoms and bonds in the model, individual strands could get mangled. Waiting for the structure to untangle on its own would cost significant time. Realignment of the helices negates the benefits of using high-quality prediction for generating the initial model. However, the elastic network acts as a safety harness against these distortions. The stiffness of the complete network ensures that the structure retains its original form. This safety feature becomes obsolete once the cascade reaches the high-resolution domain again. The restraints are now sequentially released from the *enrgMD* elastic network. Inter-helical elastic bonds, which enforce the helical lattice and backbone repulsion, are removed first, at a resolution around 15 Å. At this point, individual helices are distinguishable in the electron density data. Consequently, the elastic network's functionality to mimic the long-range electrostatics of the solvent is effectively replaced by the grid-based potential. Thereby, the larger simulation time step can be kept. At a resolution below 10 Å the backbone and helical grooving becomes discernible. Intra-helical bonds, which previously enforced the internal structure of the helix, are no longer needed and therefore are discarded. Only the hydrogen-bonding part of *enrgMD*, stabilizing base-pairing, remains present throughout the procedure. Without this base-pair network, the structures would collapse to the center of the helix as the grid-based potential exceeds the non-bonded interaction of the force-field. These elastic springs might only be removed once experimental resolution reaches 5 Å for the whole structure and all individual nucleotides become observable.

The model building protocol is completed with an energy minimization phase. The phase consists of gradient descent to equilibrate the structure to the intrinsic potential.[130] During this step, the strength of the grid-based potential is increased compared to previous stages.
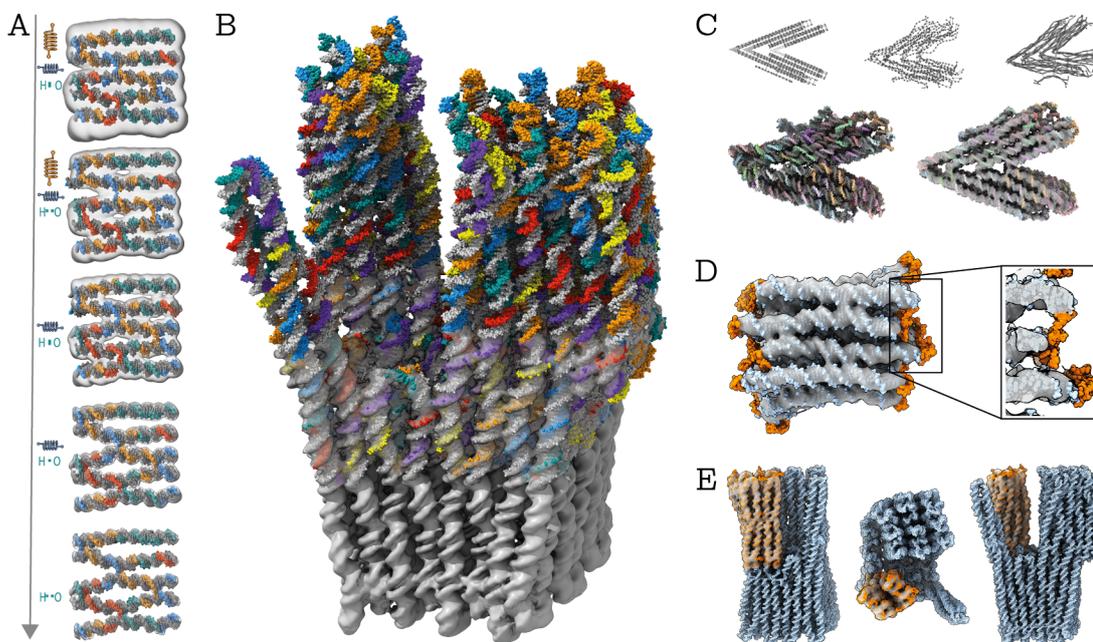
Figure 2.1 | Shrink-Wrap Protocol: (A) *Schematic Shrink-wrap model building illustration, top to bottom. Overlay of adapted model and current cascade step electron density data with increasing resolution. The different types of active enrgMD elastic network restraints are depicted by the symbols next to each step: inter-helical (orange, vertical springs), intra-helical (blue, horizontal springs), and hydrogen bonded base-pairs (H-O).* **(B)** *Exemplary overlay of the map and corrected pseudo-atomic model for the Twisttower object.* **(C)** *Seamless integration of structure prediction into the workflow, illustrated with the multidomain Corner Study Version 3 object. The multi-resolution structure prediction (top row) provides the starting configuration for the shrink-wrap protocol. Therefore, the initial atomic model (bottom left) is as close as possible to the final model derived from fitting to cryo-EM data (bottom right).* **(D)** *The exclusion of single-stranded passivization poly-thymidine strands termini (red) from MDff. The double-stranded core of the DNA nanostructure is represented at high resolution in the electron-density data (overlay). Inset shows zoom-in of 2 passivated helical termini.* **(E)** *Subset fitting of the single 4×4 domain from the Twisttower object. Selected atoms (red) are affected by the electron density data (overlay), while the remaining structure (gray) performs only a regular enrgMD simulation. Views from different directions. Panels (A-B) adopted from [32].*

A more detailed description of the protocol and the parameters used is given in Section A.2.1. Figure 2.1A depicts snapshots of the network-driven cascaded relaxation procedure. The displayed model is a small subsection of the so-called *Twisttower* object [32]. During shrink-wrap model building this procedure is performed on the complete cryo-EM data to

construct a full PAM (see Figure 2.1B).

Note: The protocol presented here is an updated and improved version of the one published in *Feigl, Kube, Kohler, et al.* [32]. Deviations from the original protocol include:

- **Initial prediction with *mrDNA*:** The Python package enables the utilization of advanced relaxation for the initial model creation. The multi-resolution prediction of *mrDNA* replaces initial models previously generated with *enrgMD*. These improved initial models also reduce the overall simulation time. Multidomain structures like corners and triangles, which otherwise would have required manual intervention in form of correctly placing the domains relative to each other, can be processed fully automatically. Figure 2.1C shows snapshots of the initial model building and fitting for the *Corner Study Version 3* object.

- **Simulated annealing:** Inclusion of this technique [106] makes the protocol more robust and helps resolve local minima automatically.

- **Optional exclusion:** The scheme of MDff supports the exclusion of groups of atoms. We recommend the omission of single-stranded passivization loops from the MDff force-field due to their insufficient resolution in the electron density data (Figure 2.1D). Complete removal from the PAM is not recommended as it might change the topology of the model. Figure 2.1E illustrates how the same methodology can be utilized to fit subdomain data generated via multi-body processing without removing contextual topology.

The addition of *mrDNA* requires GPU-accelerated software, increasing the hardware requirements. Structure prediction using this framework takes a couple of hours [116] on a standard GPU. The PAM building is performed in NAMD2, making it compatible with GPU acceleration. Using our protocol, we can build quasi-atomic models for DNA origami nanostructures of up to 10,000 base-pairs in less than half a day on a regular desktop computer. The protocol runs fully automatically, with the only exception being the rigid-body docking which has to be done at the beginning of the fitting protocol or after the *mrDNA* structure prediction. Rigid body docking takes only a couple of minutes, is not hardware intensive, and can be performed on almost any device.

2.1.2   Validation

To assess the quality of the final pseudo-atomic model we use map-to-model cross-correlation coefficients (see Section 1.3). First, an artificial electron density map is generated from the atomic model, which is then compared to the experimental data. The significance of these correlation coefficients can be increased by selecting only a meaningful subset of voxels.

A meaningful subset can be generated by defining the minimal enclosing box, cropping at an iso-volume threshold, or in the context of the fitted PAM. Selection of the map subset via the atomic model is usually referred to as masked cross-correlation coefficient (mCCC) [156]. This approach is well suited for DNA origami nanostructures because the experimental data are well-defined [157]. Due to the designed routing and sequence of the DNA strands, all molecules present in the sample are identified, and their approximate location within the assembly is known. For masking, the experimental data are cropped to include only voxels occupied by the atomic model for the calculation of the cross-correlation. Utilizing masking removes effects on the correlation of nonzero background signal and noise in the experimental data, making it independent of the box size. All mCCC data presented here was masked using the *FitViewer* tool introduced in the next chapter in Section 3.2. Additionally, we used cross-validation via Fourier shell correlation [149, 157, 168] with the experimental volumetric dataset to check for overfitting (Appendix Section B.1).

Since the *Pointer Version 1* structure published by Bai et al.[33] has so far served as the standard reference for quasi-atomic models of DNA origami nanostructures, we remodeled it using our workflow. Our model matched the previously reported structure with a root-mean-square deviation of 6.03 Å, which is significantly lower than the reported cryo-EM resolution. In terms of mCCC, our model also provides a slight improvement from 0.894 to 0.916 compared to the published version (PDB 2YMI). Manual construction of the structure reportedly took several weeks, highlighting the improvement our semi-automated approach provides.

2.2   Model Building Framework *dnaFit*

Executing the full model building protocol and subsequent analysis require experience in MD simulation with NAMD2 and basic programming skills. These requirements pose an entry barrier, especially compared to protein model building, where applications with extensive functionality and intuitive graphical user interfaces exist. To this end, we have developed the DNA origami model building and analysis framework *dnaFit* that is designed to cover the complete PAM building and analysis pipeline. The tool is implemented in Python 3, easy to install via the standard Python package pipeline, and distributed as a command-line interface (CLI) with ample documentation and additional resources. It handles not only preparation with *mrDNA* and the fitting procedure automatically but also allows direct conversion of the output data for upload to the PDB, serves as the backend for the analysis tool *FitViewer* (see Section 3.2), and facilitates analysis of the quasi-atomic model.

2.2.1   Availability and Installation

The package is publicly developed on GitHub at `https://github.com/elija-feigl/dnaFit` under the GNU General Public License 3. Development as an open-source Python package also guarantees seamless integration of improvements and interdependence with other packages like *mrDNA*. This package is still under active development and therefore subject to change. Instructions given below reflect package version *1.0.1.alpha*. Please refer to the official documentation to get the most up-to-date information.

To use *dnaFit* download/clone the repository and install it with a standard Python installation frontend. Detailed, up-to-date instructions are available on the project's README page (see URL above). As *dnaFit* acts as an interface for both *mrDNA* and NAMD2. Both these programs have to be installed to use the full functionality of the package.

An extensively documented command-line interface allows the user to browse, select and execute the various functions of the application layer of *dnaFit*. The CLI is available upon installation of the Python package. For more detail on software availability see Section A.2. A detailed description of the package structure and core routines is presented in Section A.3.

2.2.2   Functionalities:

Implementing the cascaded *mrDNA*-driven MD flexible fitting script module, the CLI exposes several commands to the user. The list of commands and additional information is also available via the CLI help pages. We first list currently available tools before presenting an exemplary minimal workflow to illustrate the building process of a pseudo-atomic DNA origami nanostructure with this package.

Main Workflow: The following three commands represent the complete preparation and model-building workflow.

1. **mrDNA:** Starts a *mrDNA* structure prediction with custom settings optimized for initial models of lattice-based DNA origami nanostructures, including adapted settings for multidomain designs. A caDNAno design file, the scaffold strand sequence, and the experimental cryo-EM data have to be provided. At the end of the *mrDNA* run, several processing steps are performed. These steps include the center of mass docking of electron density and the initial model. The encapsulation of *mrDNA* and the processing of the prediction results within one package are intended to simplify both the initial model generation and the rigid-body docking. Experienced users

might elect to create the initial model directly with *mrDNA* or *enrgMD* instead of using this encapsulation. This step takes 1 to 4 hours on standard hardware, depending on the size of the structure.

2. **vmd_info:** We suggest using VMD for rigid-body docking. Executing this command prints instructions, supplementary information, and all required commands for rigid-body docking with VMD.

3. **fit:** Performs the complete shrink-wrap fitting protocol as described in Section 2.1. All relevant parameters and settings are generated automatically. A fully automatic routine generates the cascade data by internally calling VMD, before successively executing each step. Progress is monitored and passed to the user via a logging interface throughout the protocol execution. Upon finalization of the protocol, a final pseudo-atomic model and a masked volumetric dataset of the electron density data are generated.

Additional Features: In addition to the core protocol, we present a list of features and tools implemented with the package. All items listed here are also already part of the main workflow described above. However, they either expand the functionality presented in Section 2.1 or are exposed as additional commands to allow their use in customized workflows.

- **Protein Data Bank compatibility** The initial PDB format (.pdb) has become too restrictive to accommodate structures of the size of DNA origami nanostructures. A common alternative, which presents more flexibility and can accommodate more metadata, is the Macromolecular Crystallographic Information File (mmCIF) format (.cif) [169]. This format is dictionary-like and unlimited in terms of the structure's size and compatible with the Protein Data Bank. To facilitate deposition of the pseudo-atomic models to the online database we implemented a conversion tool from *pdb* to *cif* format, specific for DNA nanostructures. The tool has since been moved out of *dnaFit* and is developed separately as *pdb2cif*, available at GitHub `https://github.com/elija-feigl/pdb2cif`. The *dnaFit* package provides a short wrapper command for *pdb2cif*.
  All quasi-atomic models generated with *dnaFit* are also converted to mmCIF and can therefore directly be uploaded to the PDB without the need for additional conversion tools. As the current map resolution is not good enough to properly resolve their position, the removal of Hydrogen atoms (available as an additional option in *pdb2cif*) is recommended to reduce the file size.
- **Masking**: *dnaFit* also implements functionalities to manipulate the volumetric data and create masks using an atomic model. The masking command provided by the

package generates the voxel envelope of the volumetric dataset and PAM required
for the quality assessment by the mCCC metric.

- **Linkage generation**: The package not only provides the linkage for caDNAno
  design, PAM, and cryo-EM data internally but can also produce human-readable
  output. The linkage is also required for the *FitViewer* tool presented in Section 3.2.
- **Low-resolution mode**: For maps with an overall resolution over 14 Å we present
  an alternative protocol. In contrast to the high-resolution setting, the intra-helical
  enrgMD elastic network is never removed. The helical structure is enforced throughout
  the fitting process.
- **Subdomain mode**: The method of MDff provides control over which atoms are
  affected by the grid-based potential. For the default workflow, this functionality
  excludes single-stranded passivization strands at the edge of the structure. These
  strands have a significantly worse resolution due to their mobility. The same approach
  can fit incomplete electron density data or subdomain data while maintaining the
  context of the full atomic structure. A file defining the subset to be fitted into the
  data can be generated using the *FitViewer* and passed as an optional argument to
  the *fit* command.

2.3   Pseudo-Atomic Models of DNA Nanostructures

Before the first publication of our protocol (Feb 2020) [32] approximately 65,000 base-pair
coordinates of DNA-only structures had been deposited at the PDB. We have since released
around 130,000 base-pair coordinates to the PDB, effectively doubling this metric. The
coordinates stem from eight pseudo-atomic models that have been fitted into high-resolution
cryo-EM data (see Figures 2.2- 2.4). These nanostructures include objects for high precision
twist design (see *Twisttower* variants Figure 2.2A-B) and a remodeled *Pointer Version
1* [33] and its variant *Pointer Version 2* [32] (see Figure 2.2C-D).

The capability of our method to create PAMs for DNA nanostructure of varying sizes
is illustrated by two examples in Figure 2.3. The *16-Helix Bundle* (Figure 2.3A) with
1033 scaffold bases is a full order of magnitude smaller than the dual scaffold *126-Helix
Bundle* object Figure 2.3B. While capable of building pseudo-atomic models of all sizes,
the *126-Helix Bundle* took around 72 hours to complete, as contrasted to only around
three hours for the *16-Helix Bundle*.

In Figure 2.4 three variants of brick-like structures are displayed. The designs *A-Brick* and
B-Brick (Figure 2.4A-B) can dimerize using shape complementary recesses and protrusions
on their sides. These pseudo-atomic models were published in Bertosin et al.[16] and are
discussed in more detail in Section 4.1.
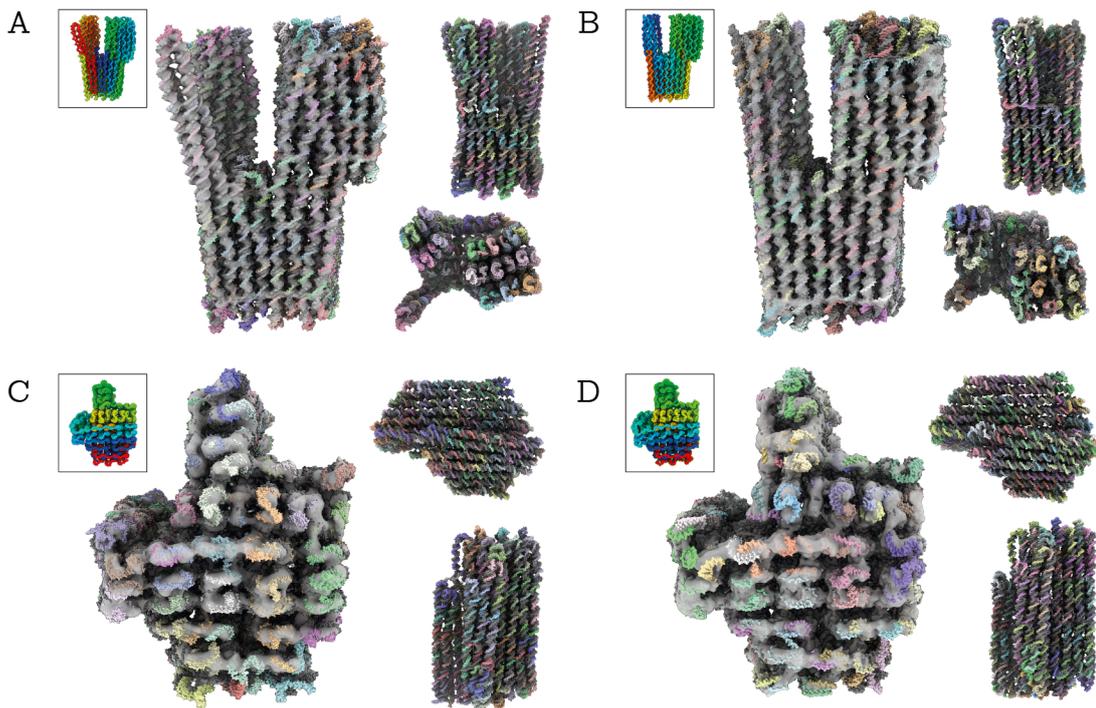
Figure 2.2 | *Pseudo-atomic model: high precision.* (A) *Twisttower Native Twist.* **(B)** *Twisttower Corrected Twist.* **(C)** *Pointer Version 1 (remodeled from [33]).* **(D)** *Pointer Version 2. Scaffold (gray), staple (colored). Inserts in the small box show only the scaffold strand colored by base ID (rainbow).*
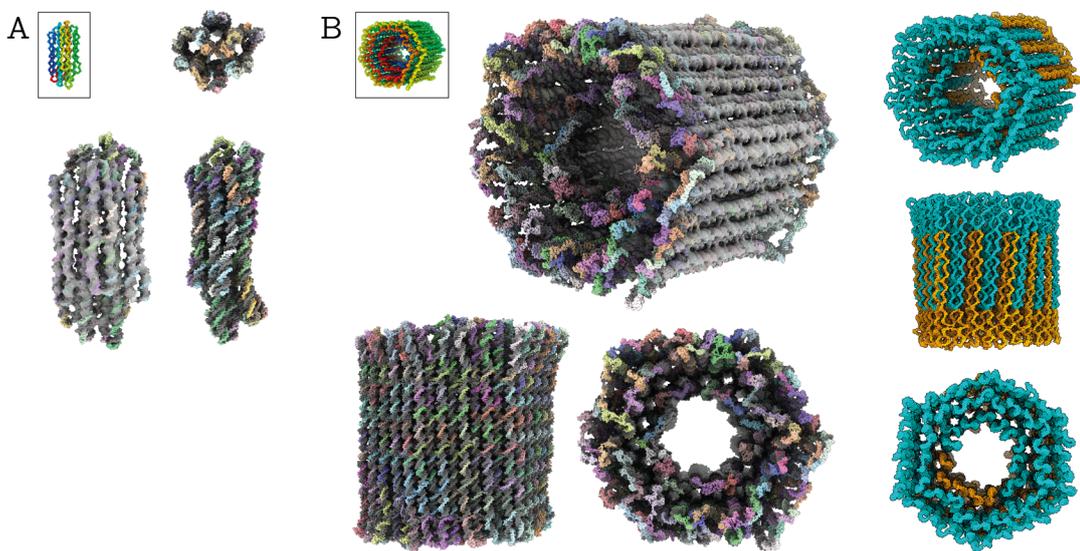
Figure 2.3 | *Pseudo-atomic model: varying size.* (A) *126-Helix Bundle.* **(B)** *16-Helix Bundle. Scaffold (gray), staple (colored). Inserts in the small box show only the scaffold strand colored by base ID (rainbow).*
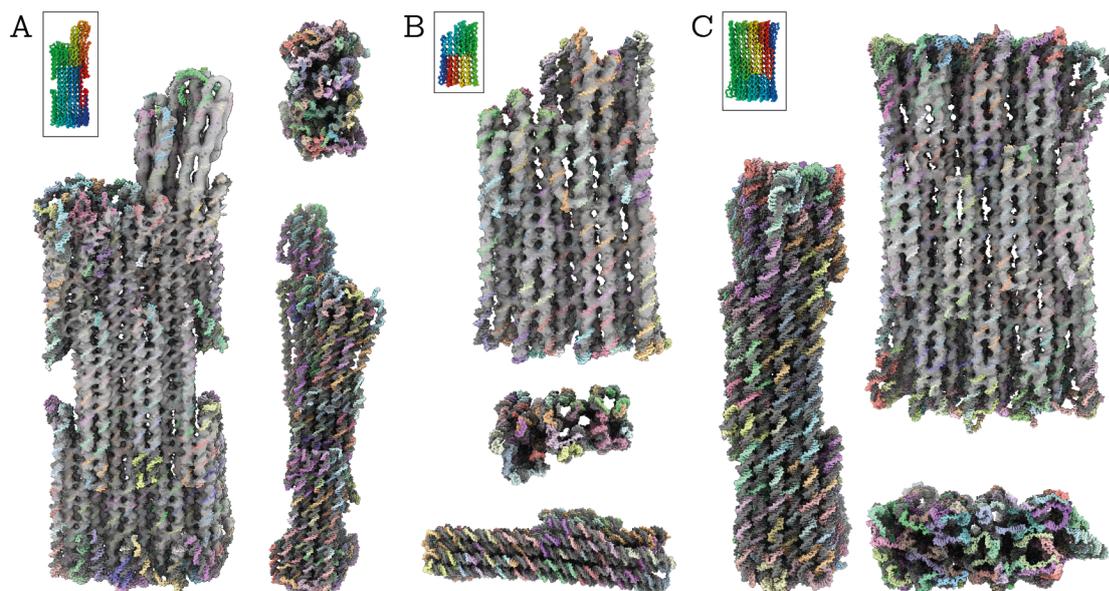
Figure 2.4 | *Pseudo-atomic model: brick-like (A)* *A-Brick. **(B)** B-Brick. **(C)** 48-Helix Brick. Scaffold (gray), staple (colored). Inserts in the small box show only the scaffold strand colored by base ID (rainbow).*

In addition to the structures published at the PDB [16, 32], we present a collection of atomic models created to support the design process for target functionalities of other research projects. These include design of specific corner angles (Figure 2.5A-B), location specific placement of moieties (Figure 2.5C-F) and validation of dihedral angles of tubule-forming triangles [170] (Figure 2.5G-H). To highlight the utility of PAMs, a selection of models displayed in this section are discussed in more detail in Chapter 4. Overall resolution data, map-model correlation values and PDB identifiers (published structures only) are listed in Table B.1 of the Appendix.
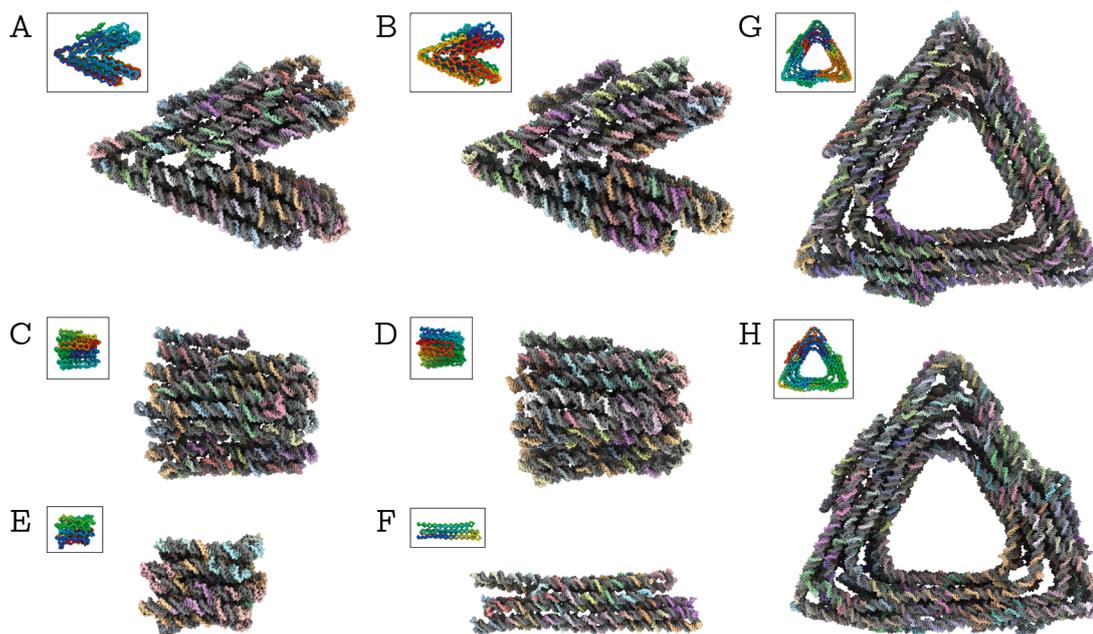
Figure 2.5 | *Pseudo-atomic model: unpublished overview (A-B) Two variants of a 45°
corner objects: Corner Study Version 3* and *Corner Study Version 6.* ***(C-D)** Two variants of
the high-resolution structure determination platform for functional RNA and DNA: Whopper* and
Whopper Version 2 ***(E)** Smaller variant (1033 bases) of the Whopper imaging platform: Whopper
Junior.* ***(F)** Triplex Bullet for pH-induced structural changes via DNA triplex crossover formation.*
***(G-H)** Low-resolution models of tubule-forming A-Triangle and B-Triangle [170]. Inserts in the
small box show only the scaffold strand colored by base ID (rainbow).*

2.4 Conclusion

Shrink-Wrap Fitting enables the generation of pseudo-atomic models for DNA origami
nanostructures in a semi-automated fashion within less than 24 hours. This presents a
significant speed-up to the previously reported [33] multi-day process that also required
several stages of manual intervention by an expert. As has already been the case for
proteins [36, 157], pseudo-atomic model construction can now be incorporated into the
standard analysis workflow of high-resolution imaging of DNA origami nanostructures.

With *dnaFit* we presented a model-building framework that makes this protocol easily
accessible for the field. Since its original publication [32], the methodology has seen constant
improvements and is now also capable of automatically fitting multidomain structures by
utilizing the structure prediction tool *mrDNA* [116]. The Python package is extensively
documented, guiding the user through the fitting procedure and providing final data, that
are readily prepared for upload to the Protein Data Bank.

So far we have published eight atomic models with the data bank, comprising approximately
130,000 base-pair coordinates of electron-density data with a resolution of 10 Å or better.

In the upcoming years, the resolution of cryo-EM data of DNA origami nano-objects is expected to improve due to the constant advances in protocols and equipment. Consequently, the importance of building pseudo-atomic models will increase.

In the future, virtual reality [171, 172] might provide a more native representation of volumetric data. However, the annotation and topological context provided by the coordinates and covalent bonds of the pseudo-atomic models will remain crucial for a high-precision evaluation of the data.

Chapter 3

Context-Based Zoning & Masking

The *dnaFit* package presented in the previous Chapter 2.2 not only implements the model building protocol but also serves as the base for annotation and categorization of the volumetric cryo-EM data. At its core, this functionality is creating a linkage between the three formats present in the design pipeline: caDNAno design file, experimental volumetric electron density, and coordinates of the pseudo-atomic model. Here we present a browser-based Jupyter notebook [173] called *FitViewer* that uses this linkage to help analyze the experimental dataset. The notebook allows its user to interactively select and zone in on a specific region or design motif. It can also create a PAM of the subset and crop the volumetric data to contain only those voxels relevant for the selected bases. The *dnaFit* package serves as the main resource for this slicing and zoning tool. With the package already installed, no further preparations are necessary, shortening the installation procedure and increasing usability.

3.1   Context Based Model to Design Linkage

The representation of volumetric datasets as an iso-surface [174] is a visualization technique commonly used in electron microscopy. Here, a surface containing all points of constant value in the volume is drawn. For cryogenic electron density data, this iso-surface corresponds to the nuclear envelope of the DNA helices if the appropriate threshold value is chosen. These surfaces are a capable tool to assess the nano-object's overall shape. But, if used as the only tool, they are limited in their use for detailed investigation of a structure. Average multilayered DNA origami consists of around 40 helices, approximately 200 base-pairs long [26, 86]. Consequently, about half the helices are on the interior and partially hidden by the iso-surface of the outer helices.

To properly inspect these helices, the volumetric dataset has to be manipulated to remove obstructing helices from view. Typical techniques for this are planar slicing or manually deleting parts of the data, for instance, the outer helices. These procedures are cumbersome to use and time-consuming, which makes them inefficient analysis tools. First, planar slicing — where the dataset is cut off along a user-defined plan — cannot capture the non-planarity of layers in DNA nanostructures [32]. These discrepancies from the design framework are caused by the flexibility and deviations of the actual geometry of B-DNA from the idealized DNA origami lattices [33]. Consequently, planar slicing along one layer will always clip some selected helices, while neighboring helices might still be partially visible. For twisted and curved objects, this shortcoming is particularly detrimental. Additionally, the orientation of the slices might be restricted to the axis of the volumetric grid, not allowing arbitrarily chosen planes. Secondly, manual manipulation, often performed using a spherical eraser tool within the visualization software (f.i. in Chimera [175]), is a slow and imprecise technique. Depth perception in 2D renderings of volumetric data is limited and highly dependent on the rendering technique. To properly place the eraser tool, the 3D object has to be moved and rotated constantly throughout the process. If high precision is required for the cutting, multiple erase operations with varying sphere radii are necessary. This same technique, which might take several hours if multiple masks are required, is also often used to create masks for focussed refinement of electron density data, for example with multi-body-reconstruction [99]. Finally, these approaches do not exploit the inherent connection between the experimental data and its design. Figuring out which part of the design is represented by which subset of voxels is left to the user. Thereby experience and expertise are required while the workflow remains inherently imprecise. Especially for larger structures, navigation of the volumetric dataset is made difficult by the repetitive nature of the lattice-based design. Another common strategy for high-resolution maps is crossover-counting. By starting from a well-resolved structural motif, the design can be traversed by counting the number of crossovers between the helical pair. Positions between crossovers are then estimated based on the number of bases passed. However, the incomplete assembly of parts of the structure and variation in the local resolution of subdomains makes this strategy error-prone.

We tackle these challenges by fully automatizing these zoning & slicing tasks and completely removing manual data manipulation. Automatization is achieved by using the context assigned to the data by the PAM. The selection of subdomains is no longer achieved by manipulating the volumetric data but by using the language of the design process: caDNAno helix number and base position. Translation from the design space to 3D volume space is performed by linking the voxel index to the atomic index and hence the coordinates

of the atoms in the fitted model. This conceptual change allows the user to directly contextualize a design question with the experimental data represented by the volumetric data set.
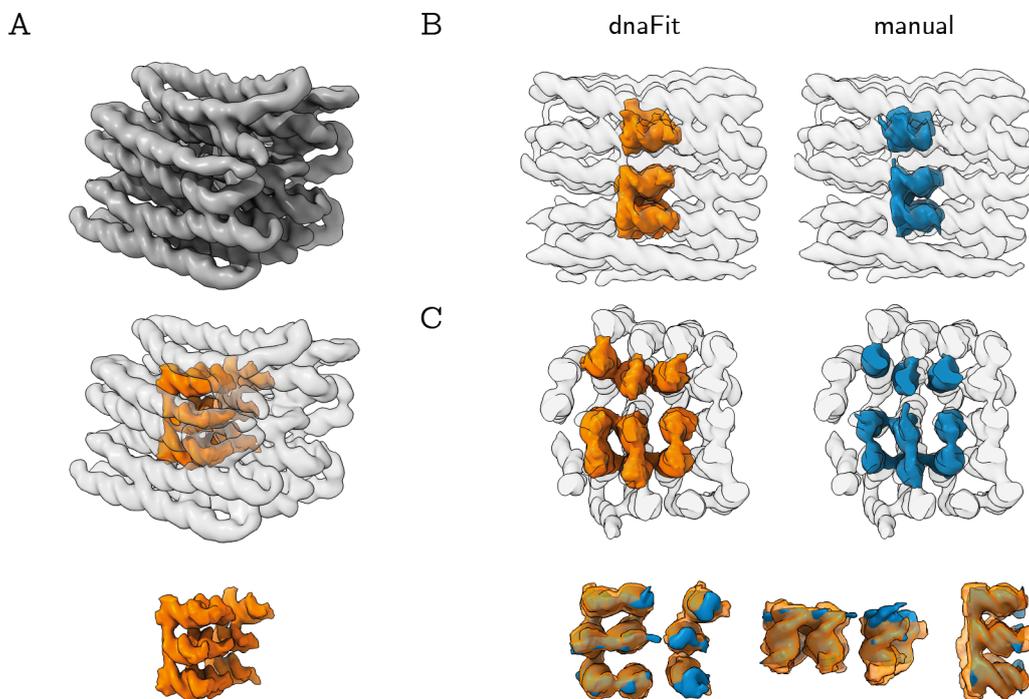
Figure 3.1 | **Subset Zoning: (A)** Subpart of a seven base-pair long segment of the interior 3×3 helices (orange) of the -JrV3 object selected via model-based linkage. **(B)** Overlay of full structure with automatically (orange) and manually (blue) selected subdomain and **(C)** their direct comparison.

Thereby, a set of voxels can be selected based on a specific site or motif in the design. This annotation of the three-dimensional continuous dataset makes design-specific data analysis possible. It facilitates the creation of a pseudo-atomic model of the selected subset and cropping the volumetric data to only those voxels relevant for the chosen nucleobases.

To illustrate the effectiveness of this approach, we selected a seven base-pair long segment of nine interior helices of the *Whopper Junior* object that spans two crossovers motifs using the Jupyter notebook presented in the next Section 3.2. The subset of this segment consisting of 63 base-pairs is displayed in Figure 3.1A. Additionally, we compared this automatically-generated subset to manual zoning using the ChimeraX spherical-eraser tool (Figure 3.1B-C). Attempting to manually select individual bases at a resolution above 5 Å is difficult, which can be seen by the slight deviations between the two subsets. The eraser tool acts on the voxel data itself, with the user often visualizing the data using an
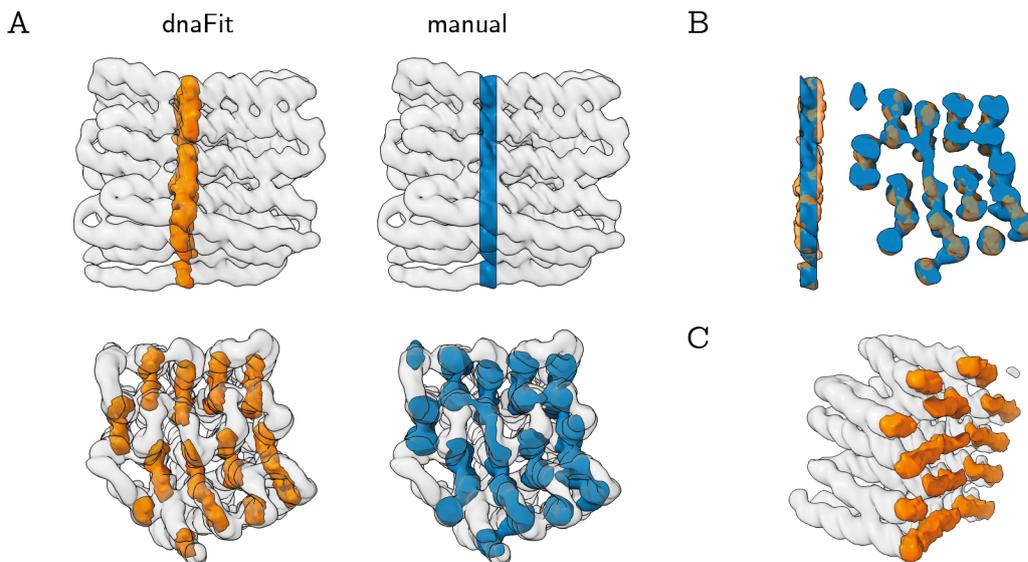
Figure 3.2 | *Layer Slicing: (A) Overlay of full structure with automatically (orange) and manually (blue) selected single base layer at helix position 32 of the Whopper Junior design. **(B)** Direct comparison of both techniques. **(C)** Linkage-based selection (orange) in the context of the rear half of the structure.*

iso-surface. This can cause decoupling of input and visual feedback and hence limits the precision. Moreover, the ChimeraX spherical-eraser tool required around 10 minutes, while the zoning via linkage only took a couple of seconds.

The previous example necessitated the use of the spherical eraser as the lattice of the structure is slightly distorted, resulting in a non-planar alignment of the helices. Hence, we compare another technique, the ChimeraX planar slicing tool, with our method. One use case for this tool is the creation and analysis of the cross-sectional slices of the structures. Figure 3.2 shows the comparison of a single base slice at base-position 32 in the *Whopper Junior* design. Planar slicing is fast and intuitive to use. Consequently, the automatic approach via quasi-atomic model-based linkage is not significantly faster than manual slicing. However, determining the exact position via manual selection can be difficult if it does not (as is the case in the chosen example) coincide with a well-resolved crossover position. Again, the direct comparison of automated and manual selection (Figure 3.2B) reveals deviations from lattice planarity. The presented example only displays a negligible effect but especially slices of objects with larger diameter (for instance the 8×8 domain of the *Twisttower* or both *Pointer* variants) can warp strongly. There, the planar slicing tool is no longer capable of adequately representing the cross-section of a DNA origami nanostructure at a predetermined base position within the lattice.

In summary, an automated approach is faster and more accurate than manual selection tools. Linkage-based selection via a fitted PAM allows specifying locations based on the context of the designed lattice. Helix ID numbers and base positions are translated to voxels utilizing the atom identities and coordinates. The accuracy of this zoning is restricted by the fitting precision alone, with at least 14 Å for high-resolution maps shown in this work. This level of precision surpasses any manual selection tools that rely on high-resolution landmarks like crossovers for orientation. Automated linkage is also independent of distortions of the ideal lattice model, like a twist or barrel-shape deformations [32] that can be observed regularly. Apart from visual zoning, the context provided by the exact coordinates enables precise quantification of the selected design subparts and motifs.

3.2 FitViewer: Zoning, Cropping & Masking

The model-building protocol in our *dnaFit* package is distributed as a command-line interface. The same approach is not suited well for the zoning workflow described in the previous section. With the *FitViewer* Jupyter notebook, we provide a simple-to-use frontend to perform slicing, zoning, and masking. Here *dnaFit* acts as a backend performing the linkage and cropping operations. Using a notebook enables the integration of interactive widgets, visual representation, and text-based elements. Within the notebook, a set of instructions and examples is given next to each executable cell. Jupyter notebooks are browser-based, making them independent of the operating system, and can be hosted on a server. At the same time, a Jupyter notebook is a fully functional Python shell making the inclusion of other libraries straightforward. Thereby, *FitViewer* might also be seen as a starting-off point to make use of the full potential of the *dnaFit* library and create and operate other or more complex tasks and analysis workflow. Notably, we have used modified versions of the notebook to compute absolute dimensions of the structure, quantify their twist and generate helical trace images. In Chapter 4 a few examples for this approach are given.

Figure 3.3 shows an overview of the base interface of the notebook. The notebook is enhanced with extensive documentation, instructions, and examples to help users execute and customize the individual steps while likewise ensuring that the full functionality of the *dnaFit* package is evident. The capability of a Jupyter notebook to combine Markdown text and executable code allows seamless integration of both aspects: functionality, and documentation. These features make this format suitable as an interactive computational environment for tasks like the linkage selection presented here. All Markdown text is removed in the figure to provide a better visual overview. The following paragraph covers a detailed description of the interface's three main parts:

```
import dnaFit.viewer.viewer as app
from pathlib import Path
```

1. Parse File-Paths & instantiate Viewer

```
# specify working directory and design name
wd = "/Users/elijafeigl/shrik-wrap/wjrv3/"
name = "wjrv3"

json = Path(wd +name + ".json")      # caDNAno design file
psf = Path(wd + name + ".psf")       # topology file
coor = Path(wd + name + ".pdb")      # coordinate or trajectory file
mrc = Path(wd + name + ".mrc")       # cryo-EM map
seq = Path(wd + name + ".seq")       # scaffold sequence file

viewer = app.Viewer(conf=coor, top=psf, mrc=mrc, json=json, seq=seq, is_mrdna=True)
```

2. Subset Selection: Pseudo-Atomic Model

Option A: Selection Widget

```
sliders = viewer.select_widget()
```



```
helix and base: ([1, 0, 3, 2, 8, 9, 6, 7, 11, 10, 13, 12, 18, 19, 16, 17, 21, 20, 23, 22], range(15, 21))
number of selected bases: 234
```

Option B: Manual Selection

```
## example Option B:
helix_selection = [1,2,3,4]
baseposition_selection = range(10,30)
atoms_selection, color_dict = viewer.select_by_helixandbase(helix_selection, baseposition_selection)
```

Option C: Fit Specific Selection

```
## example Option C: all double-stranded segments with staples beginning with the sequence ATCG
""" either use MDanalysis directly to f.i.:
        select all staples beginning with a specific sequence,
        via the intersection of all staples with the correct base condition
        start with all segments and filter with additional conditions
"""
ATCG = viewer.u.segments
for idx, X in enumerate(["ADE", "THY", "CYT", "GUA"]):
    ATCG = ATCG.segments & viewer.u.select_atoms("resname {} and resid {}".format(X, idx+1)).segments
atoms_staple  = ATCG.segments.atoms

""" or use custom selection methods to f.i.:
        pick all paired scaffold bases, while excluding Hydrogen atoms
"""
atoms_ds = viewer.select_ds_dna()
atoms_sc = viewer.select_scaffold(atoms=atoms_ds)
atoms_selection = viewer.select_without_H(atoms=atoms_sc)
```

3. Cropping & Zoning

Cropping: Electron Density Data

```
#create subset MRC-file
viewer.write_mrc(atoms_selection, name_out, context=context, cut_box=False)
```

Zoning: Atomic Model

```
#create subset pdb-file (chimeraX compatible)
viewer.write_pdb(atoms_selection, name_out, single_frame=True, frame=-1, as_cif=True)
```

Tools: Subset Fitting & Masking

```
#create grid.pdb-file for subset fitting
viewer.write_custom_gridpdb(atoms_selection, name_out)
```

```
#create binary mask
write_binary_mask(atoms_selection, name_out, context=context)
```

Figure 3.3 | *FitViewer interface: (1. Input)* Selection of file paths and initialization of the viewer object. *(2. Subset Selection)* Three different methods for subset selection. Inserts (orange boxes) in Option-A show a comparative view of the same depiction in caDNAno2. *(3. Cropping and Masking)* Code Snippets for output generation.

1. **Input:** After importing the relevant resources by executing the first cell, all input files are specified by the filesystem path. These input files are the caDNAno design file, a plain text file with the scaffold sequence, topology and coordinate file for the pseudo-atomic model, and the electron density volumetric data. These files are then parsed to create a *Viewer* object that can subsequently be used to perform the individual tasks of the workflow. Parsing of the molecular dynamics simulation data is performed using the Python package MDAnalysis [176]. For cryo-EM data and caDNAno design files, we used the packages mrcfile [177] and a customized version of nanodesign (see Section A.2.4), compatible with Python 3, respectively. Internally a linkage lookup table is created using the *dnaFit* package that assigns atoms indices of the model to the overlapping voxel coordinates of the cryo-EM data and helix and base position in the design file.

2. **Subset Selection (Slicing and Zoning):** The notebook presents several options to select subsets from the provided data. The first option is the selection of helices and base positions using a widget. The widget is modeled after the caDNAno helix view and position slider. The second option is to specify their indices manually. Alternatively, custom groups of atoms are selectable based on the PAM using MDAnalysis. For instance, the example given in the instruction text describes the steps to select all

staples beginning with the sequence *ATCG* and their associated scaffold bases. Either option to select a subset returns an instance of the MDAnalysis AtomGroup class. These AtomGroups still contain the complete topological, geometrical, and chemical information about the selected set of atoms. The user can further manipulate or combine individual groups. These groups are also instrumental for more in-depth analysis and advanced computation, for instance, distances between atoms.

3. **Cropping and Masking:** Finally, the selected group of atoms is used to generate subset data. Via the *dnaFit* packages, both a pseudo-atomic model file and an electron-density data file containing only data for the selected subset can be generated and written to a file. Voxels to be cropped from the full volumetric dataset are selected by assessing their distance from each atom. All voxels that are closer than a user-defined cutoff value are included. These cropped maps can be used for validation via cross-correlation or exported as masks for further post-processing of the experimental data. An additional flag in the output function ensures required modifications, for instance retaining the same grid as the original data.

Additionally, the *FitViewer* can be used to generate the reference files required for subdomain fitting. Atoms from the subdomain can be selected using the same options described above for the zoning.

The notebook is available as open-source under the GNU General Public License 3 on GitHub `https://github.com/elija-feigl/FitViewer`.

3.3   High Precision Design Evaluation using *FitViewer*

To further illustrate the benefits of automated zoning and slicing, we present several qualitative results generated with the *FitViewer* notebook.

Firstly, we cut the *Twisttower Native Twist* object into layers according to the horizontal helix indices of the caDNAno diagram (Figure 3.4A-B). Analysis of individual slices facilitates the investigation of the topology of the model in this layer. The global twist of the 2×2 domain is visible in panels 3.4D-E as the two helices belonging to this domain are rotated by around 90 degrees. The deformation in this domain is significantly stronger compared to the other domains. The lower-left corner of the top layer (Panel 3.4D) misses a potential crossover between the first and second helix, resulting in an outward bend of the corner. This effect is enhanced by the long unconnected segment between the second and third helix in the next layer (Figure 3.4E). A prominent structural feature occurring multiple times in both *Twisttower* objects is the crossover stack. These stacks connect up to eight helices with crossovers placed at the same base position and constrict the structure
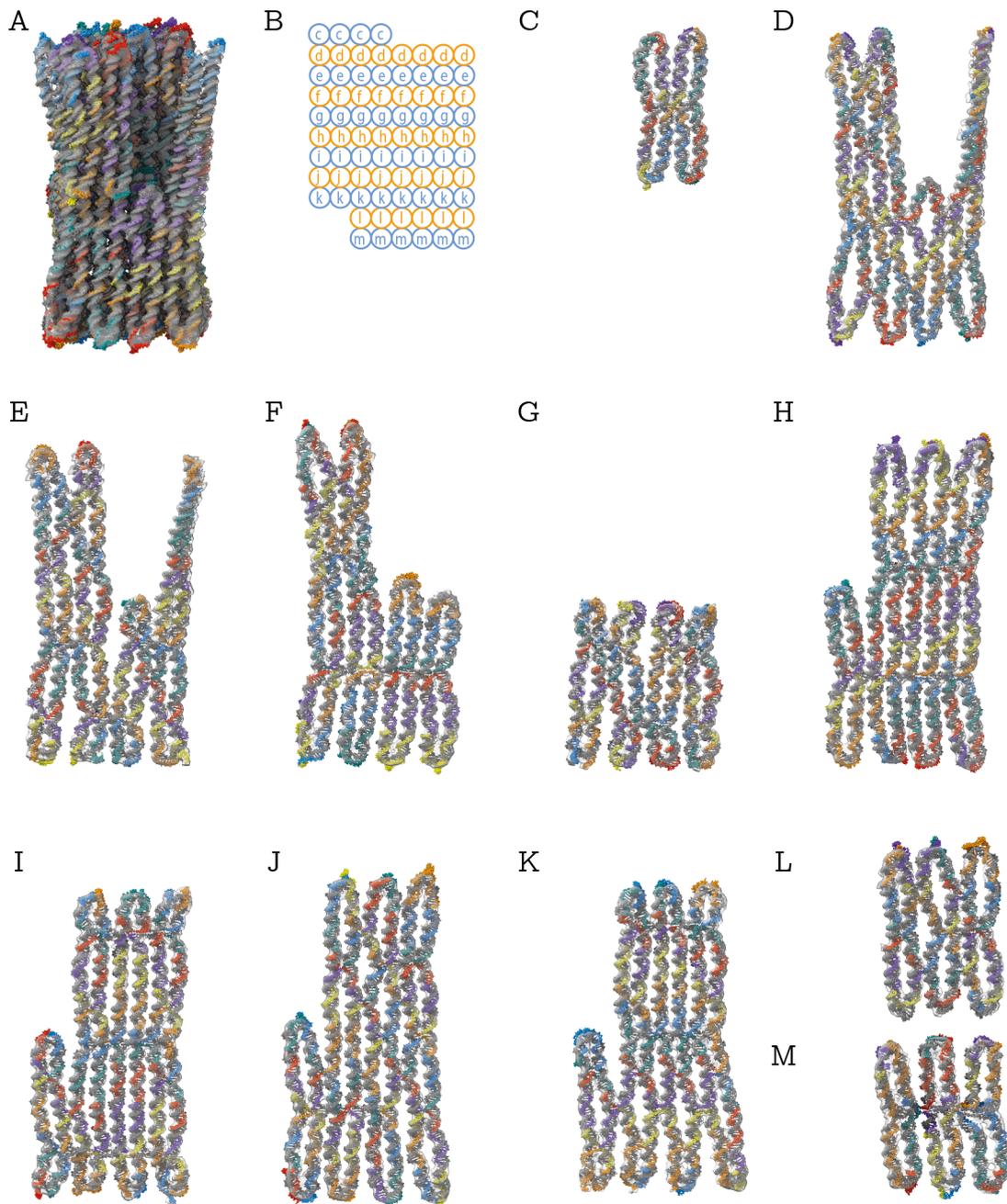
Figure 3.4 | *Layer-by-layer analysis of Twisttower Native Twist object via slicing:* *(A) Superimposition of cryo-EM data with the complete pseudo-atomic model. (B) Slice scheme indices in cross-sectional view. Labels indicate the panel in which the layer is displayed. (C-M) Layers of the cryo-EM data and model sorted by the indices in (B). Each layer is selected from the quasi-atomic model via the automated linkage and superimposed with the zoned voxel data. Adopted from [32].*

along its entire length. This effect creates an overall distortion that mimics the shape of an hourglass, observable most notably in panels (F) and (H-K). Crossover stacks are only held

together by the stacking interaction of helical ends of each Holliday junction. Hence, no covalent bond passes through them. We hypothesize that helices might shift relative to the lattice while retaining the closure of the stacks. In particular, the eight helix-wide stack on the bottom of the seventh layer 8×8 displays misalignment of the central helices below the stack compared to the remaining structure. Dehybridized base-pairs might further contribute to this artifact [79] as multiple crossovers are incidentally formed by consecutive A-T pairs. Layer analysis of the 8×8 domain also reveals that the leftmost helix pair has no crossover next to the upper edge in any of the four lower layers (panels 3.4I-K). Periodicity of the square lattice would guarantee another crossover only a few bases higher (compare six-helix-stack in 6×6 domain in panel 3.4H), but the helices are shorter than required for this connection. While vertical crossovers stabilize the structure appropriately, the helix pair is notably bent away from the main body in the last layer (K). The final layer of the 6×6 domain reveals folding defects at the central crossover stack (panel 3.4M). These defects are likely caused by stable sequence-specific secondary structures formed by the staples (see Supplement [32]). To complete the topological analysis, the procedure ought to be repeated in the vertical direction.

Secondly, we use slices by base position to describe the cross-section-dependent global twist of square lattice DNA origami structures. A superposition of these slices reveals a global right-handed twist for unmodified square-lattice DNA origami nanostructures. The twist decreases with increasing cross-section (Figure 3.5A). The causes of this twist are local torques [12] produced by the systematic under-winding of eight base helical segments that define the lattice connectivity. Global deformations can be removed by altering the number of bases in between crossovers (Figure 3.5B). Then, the effective number of bases per turn returns from a value of 10.67 dictated by the eight base increments of native square-lattice to the 10.5 bases per turn used in honeycomb-lattice. The publication by *Feigl, Kube, Kohler, et al.* [32] provides quantitative data on the global twist and instructions about its cross-section-dependent correction.

Finally, we isolate structural features of two distinct designs to highlight experimental observations of an unintentional sequence-dependent motif. During the DNA origami design process, staples are arranged to achieve two goals: maximum connectivity of the topology and optimized self-assembly [85]. To achieve this, the designer might opt for a staple break in a long segment or directly at a crossover resulting in a half crossover. Presently, sequence-dependent motifs are not considered in the design workflow. One potential arrangement arising from this set of rules is a crossover with both staples asymmetrically ending seven to eight bases after the junction. If the two staples end with a similar sequence, the motif might turn into two duplexes without a crossover. With high-resolution cryo-EM inspection guided by the annotation of quasi-atomic models, we located this

motif with varying degrees of sequence homology in two structures. In the $8{\times}8$ domain of the *Twisttower Native Twist*, the asymmetrically linked crossover can be found on top of a crossover stack. As shown in Figure 3.5C, this crossover is present in the electron density data, although at a reduced resolution compared to the other crossovers in the stack. The crossover and the topmost helix are only fully-visible at a lower iso-surface threshold compared to the surrounding structure. The sequence of the first five bases of the final crossover-binding segment is identical. Due to partial sequence homology, two different conformations are possible for staple binding. Despite the shorter matching sequence, binding in the same helix without the crossover is conceivable due to a potential energy penalty associated with establishing the crossover. The same motif is present in the *Triplex Bullet* object with a single mismatched base. Figure 3.5D depicts that the strong sequence homology prevents the crossover from assembling. Without the pseudo-atomic model, the motif is not discernible from any other subdomain without a crossover. A reduced local resolution that could result from the sequence mismatches is concealed by the lower local resolution of the helix-pair compared to neighboring helices. Not only is this pair on the exterior of the nanostructures, which reduces the local resolution due to lattice breathing and center-of-mass alignment artifacts [32]. The missing crossover also causes increased flexibility due to the decreased connectivity, further smearing its signal. A subsequence redesign with a non-homologous sequence has shown a highly resolved crossover (see Appendix Figure B.2). This series of annotated electron-density data subsets of the same design motif highlights the importance of high-precision design in DNA origami nanotechnology. We demonstrate how sequence-specific design motifs might play a role in iterative design. Ambiguous binding conformations of these structural features can either coexist in equilibrium or be determined by the degree of sequence homology in combination with the conformational change of the backbone in the Holliday junction crossover motif.
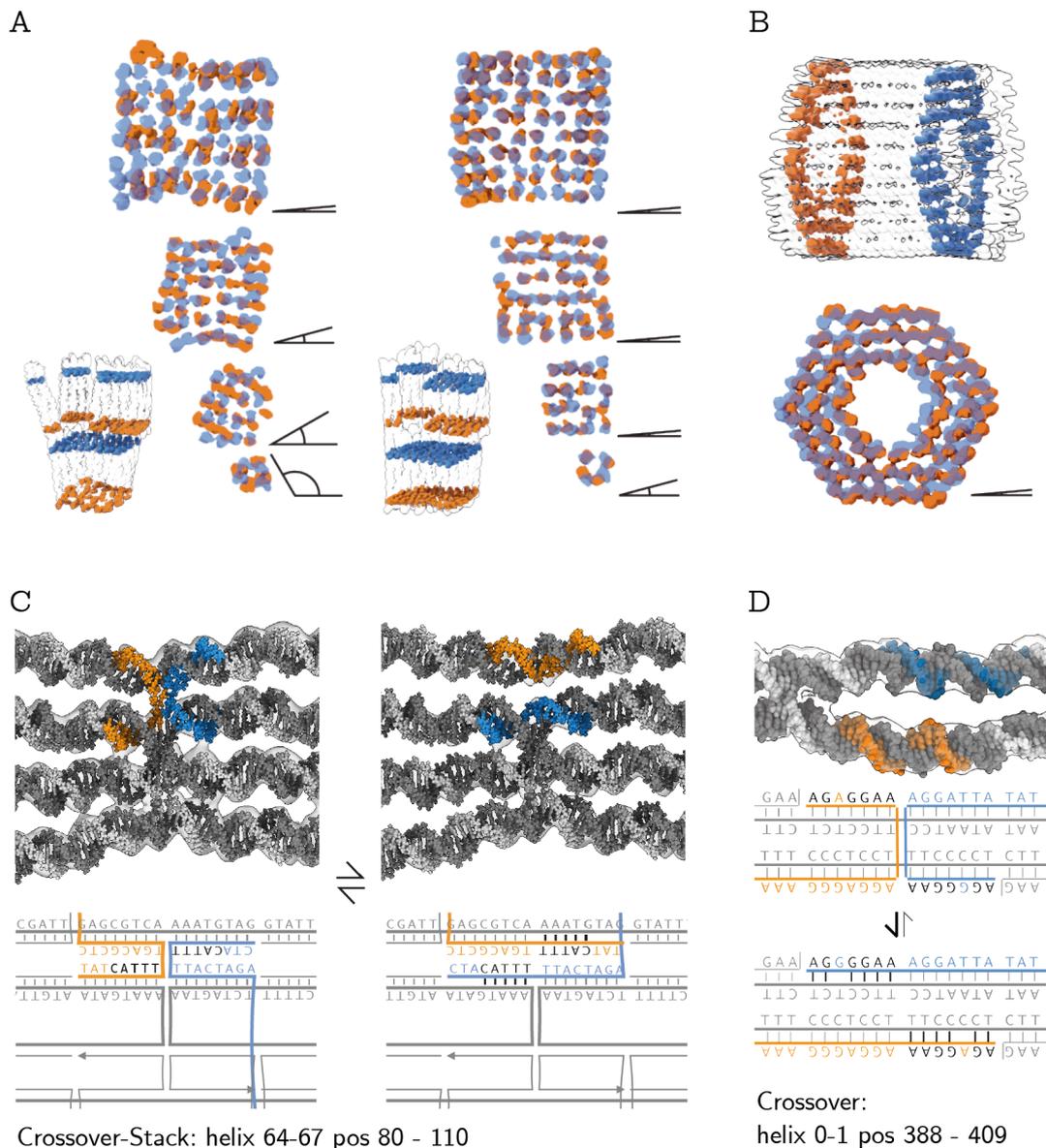
Figure 3.5 | *Qualitative details analysis of structural features retrieved via zoning.* *(A) Global twist of square lattice DNA origami of the four domains of a varying diameter of the Twisttower objects. Twisttower Native Twist (left) versus Twisttower Corrected Twist (right) designs with cross-sectional slices of the electron density data. The locations of the top (blue) and bottom slice (red) are displayed in the bottom left overlay with the full cryo-EM map. Angle tokens represent a measure for the twist angle between the two slices. (B) As in (C) for the honeycomb lattice object 126-Helix Bundle, displaying no significant global twist. (C) Twisttower Native Twist: Reduced resolution through partial sequence homology of the top crossover in a triple stack in the 8×8 subdomain. The intended crossover (left) appears equally often in the ensemble, resulting in a weaker signal than the two other crossovers in the stack. The sequence match of the first five bases creates a possible alternative conformation (right). (D) Triplex Bullet: Sequence homology of a single asymmetric crossover. As in C, but with the alternative conformation (bottom) occurring exclusively in the experimental data. Segments display a matching sequence of 6 out of 7 bases. Panels (A) and (C-D) adapted from [32].*

3.4　Conclusion

In this chapter, we introduced the interactive Jupyter notebook *FitViewer* which enables context-based zoning and slicing of the pseudo-atomic models. Geometric coordinates for the cryo-EM derived model are linked with the topological information of the DNA origami design file. Seamless integration with the *dnaFit* Python package (Section 2.2) facilitates a streamlined workflow for building and analyzing pseudo-atomic models all within the same framework. The notebook supplies the interface to perform motif-based evaluation and inspection of the experimental data based on the annotation provided by the pseudo-atomic models. By using interactive widgets these subsets can be extracted from the full volumetric dataset and analyzed independently. At the same time, the complete information of both designs and fitted models is directly available through *dnaFit*, which allows the development of specialized workflows. Inline plotting using Python libraries like Matplotlib [178] from within the Jupyter notebook enables visualization of acquired data within the same setup. Some of these more common workflows, like the creation of post-processing masks for advanced analysis of the electron density data, are directly available in the notebook.

Thus, we demonstrated the capability of *FitViewer* by slicing a multilayered DNA origami in its constituent horizontal helical layers. Cropping the full volumetric dataset of the *Twisttower* object to contain only these layers allowed a detailed evaluation of the internal distortions from the idealized lattice. Slices containing crossover stacks express a strong constriction at the position of the stack, while omitted crossovers create the opposite effect. This type of analysis also revealed structural motifs and incorrectly assembled subdomains, which were hard to detect and interpret in the complete volume data. Exemplarily, we described a sequence-dependent crossover motif present in two of the studied objects. Depending on the strength of sequence homology, the motif would, with different proportions in the ensemble, form a crossover or be present in a topological arrangement without a Holliday junction. Additionally, slicing the structures perpendicular to the helical direction yields quantifiable results for the global twist deformation of lattice-based DNA origami nanostructures [32].

Development of the source code in the public domain invites contributions from the field of DNA nanotechnology and high-resolution electron-microscopy. *dnaFit* and *FitViewer* are intended to serve as a platform for this development which hopefully includes the creation and implementation of standard analysis tools and pipelines for nucleotide-level analysis and precision design of the DNA origami nanotechnology.

Chapter 4

Quantified Structural Analysis

Geometric data obtained from the models have proven a valuable resource for quantifying and comparing structural motifs. The contextual enrichment of the experimental data by the pseudo-atomic model (PAM) allows precise evaluation of the complex information within the volumetric dataset. Here, we use the PAMs to investigate the effect of different environments on the overall shape of DNA origami structures. We illustrate how these models can be utilized effectively during the iterative design of novel nano-objects by localizing areas that require structural improvement. The atomic coordinates yield the context necessary for computing a target observable, either to assess whether a design goal was reached or to define and adapt design plans. To this end, we demonstrate the usage of the model as not only a visual inspection tool but also for the precision placement of design motifs and non-DNA moieties. Finally, we use the atomic coordinates provided by the PAM, to compute the duplex step properties of B-DNA.

4.1 Effects on Overall Shape from Coating and Buffer Changes

Remark: The work presented in this section was published in Bertosin, Stömmer, et al. Cryo-Electron Microscopy and Mass Analysis of Oligolysine-Coated DNA Nanostructures (2021) [16]. Bertosin, E. performed cryo-EM measurements and data processing of the volumetric data.

To achieve self-assembly of DNA nanostructures, electrostatic repulsion of the negative charge of the DNA backbone phosphates requires shielding. For multilayer DNA origami, aqueous solutions with monovalent or divalent ions are used to screen the destabilizing

electrostatic interactions.[26, 85, 179] Consequently, these nanostructures are prone to electrostatic-induced disassembly at non-optimal salt concentrations. DNA objects are also affected by nuclease digestion in serum, blood, or other physiological fluids. For biomedical *in vivo* applications this causes a decrease of the pharmacokinetic bioavailability of the DNA nanostructures. However, these effects can be reduced by coating the DNA with cationic cofactors [15, 89] and impeding nuclease activity by lowering the accessible DNA surfaces, especially at strand termini.

Here, two shape-complementary [10] structures, *A-Brick* and *B-Brick* (see Figure 2.4A-B), are presented that have been stabilized post-assembly. The structures are stabilized using oligolysine or polyethylene glycol (PEG)-oligolysine coating [15]. The oligolysine screens the negatively charged phosphates and decreases nuclease digestion, while the PEG prevents increased aggregation caused by the coating. We demonstrated, that the blunt-ended double-helical interfaces of the two shape-complementary structures remained accessible. However, we observed deformations of the global shape of the nano-objects induced by the coating. We used helical trace information of low-resolution quasi-atomic models to investigate these structural changes and the effect of high-density coating on the internal arrangement of the honeycomb-lattice.

Both structures were prepared at different coating ratios of oligolysine or PEG-oligolysine. Uncoated structures were also analyzed as control samples. Coating ratios ranged from 0.2:1 (N/P) (nitrogen in lysine to phosphorus in DNA) to 100:1. Only samples coated with PEG-oligolysine were exposed to physiological conditions. Structures without this coating are not stable under these conditions and were hence excluded from this part of the analysis. These origami objects were then imaged using cryogenic electron microscopy. The overall resolution of 10-12 Å for the uncoated structure is high enough to perform the full shrink-wrap protocol reported in Chapter 2. For coated structures, the overall resolution was significantly reduced, despite the same conditions and a similar amount of individual particle images. Helical grooving could no longer be observed for these structures. The decrease of discernible details is presumably caused by the cationic coating. Oligolysine and PEG attach predominantly in the grooves of helical DNA, spanning multiple base-pairs. For each DNA particle image, the positions of the coating molecules are unique, smearing their signal. This contribution to the overall intensity covers the helices, occluding high-resolution features. Nonetheless, the overall shape and topological completeness are still discernible by crossover locations. While these circumstances prohibit high-resolution model building, the data are suited for the low-resolution mode of shrink-wrap introduced in Section 2.2. There, intra-helical bonds restrain the B-DNA shape, including major and minor grooving, throughout the protocol. Since the electron density in the cryo-EM data

is highest in the helical center, the atoms will properly converge to the correct position. This is true even with the signal obscured by the coating. While the PAM created with this modified procedure is not suitable for nucleotide resolution analysis, global helical properties are accurately represented. Consequently, positional data of the helical trace, as well as overall model dimensions, can be used to quantify the experimental data.

We constructed pseudo-atomic high-resolution models for the two uncoated structures and six low-resolution models for the *B-Brick* design with different coatings and buffer conditions. The overall map resolution and mCCC from all models are listed in Appendix Table B.2. All seven models of the *B-Brick* are displayed in Figure 4.1A-G. The atomic coordinates are used to investigate the different structures independent of the cryo-EM resolution and noise of the cofactors.
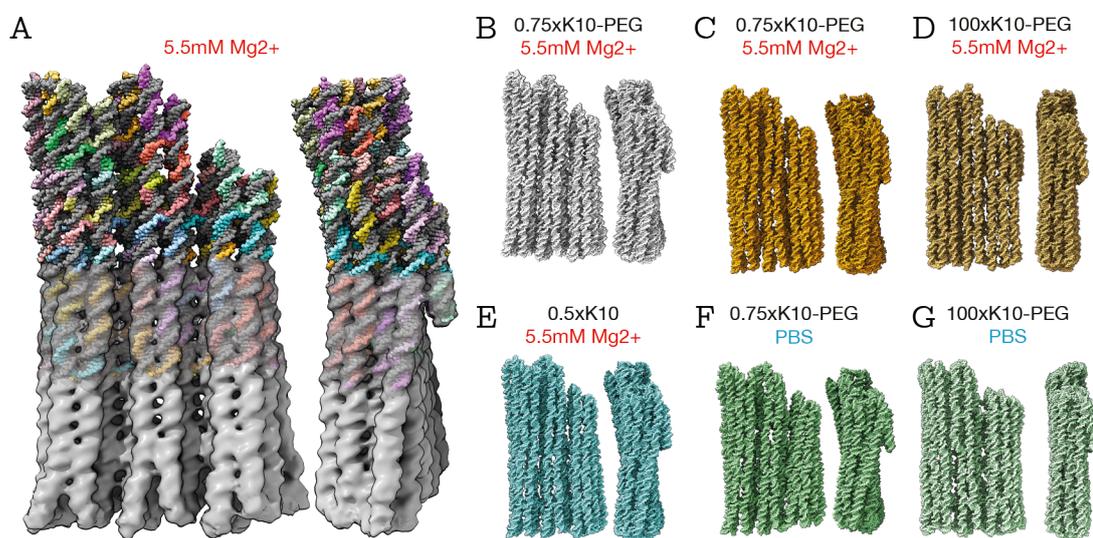
Figure 4.1 | *B-Brick Pseudo-Atomic Models*. (A) *Overlay of the cryo-EM data and pseudo-atomic model of the uncoated B-Brick in* 5.5 mM *MgCl₂. The scaffold strand is depicted in gray, staples are in color. Views from the top (left) and left side (right).* **(B-G)** *Single color quasi-atomic models for different coating ratios and buffer conditions, generated with low-resolution cryo-EM data.* **(B)** *0.2:1 N/P PEG-oligolysine coating (gray) in* 5.5 mM *MgCl₂.* **(C)** *0.75:1 N/P PEG-oligolysine coating (orange) in* 5.5 mM *MgCl₂.* **(D)** *100:1 N/P PEG-oligolysine coating (light orange) in* 5.5 mM *MgCl₂.* **(E)** *0.5:1 N/P K10-oligolysine coating (light blue) in* 5.5 mM *MgCl₂.* **(F)** *0.75:1 N/P PEG-oligolysine coating (green) in phosphate-buffered saline.* **(G)** *100:1 N/P PEG-oligolysine coating (light green) in phosphate-buffered saline. Adopted from [16].*

Selection of the central slice (4.2A) was performed using the *FitViewer* tool introduced in Section 3.2. Due to the even number of lattice columns in the *B-Brick*, the uncoated structure exhibits a slight bend in the column direction. This deformation is more pronounced for thin coatings but not visible for an N/P ratio of 100:1 in both 5.5 mM

MgCl$_2$ and phosphate-buffered saline (PBS). For these thick coatings, the PAMs reveal a transition of the internal structure of the lattice. Helices are shifted to form a compact square-lattice-like arrangement with a three×eight cross-section. The connectivity of individual helices in this deformed lattice is shown in Figure 4.2B. The bending deformation observed for thin coatings likely constitutes a beginning transition to the denser helical arrangement. The global twist of the object also disappeared for high-density coatings correlating with the lattice reorganization. However, a similarly distinct deformation is not possible for any cross-section. In contrast to the *B-Brick*, the *A-Brick* design has a larger cross-section and an odd number of lattice rows. Therefore, it cannot collapse into a single compact helix rearrangement but only deforms partially. Its low-resolution electron density data are a superposition of different helical alignments, wherefore a similar analysis is not possible [16].
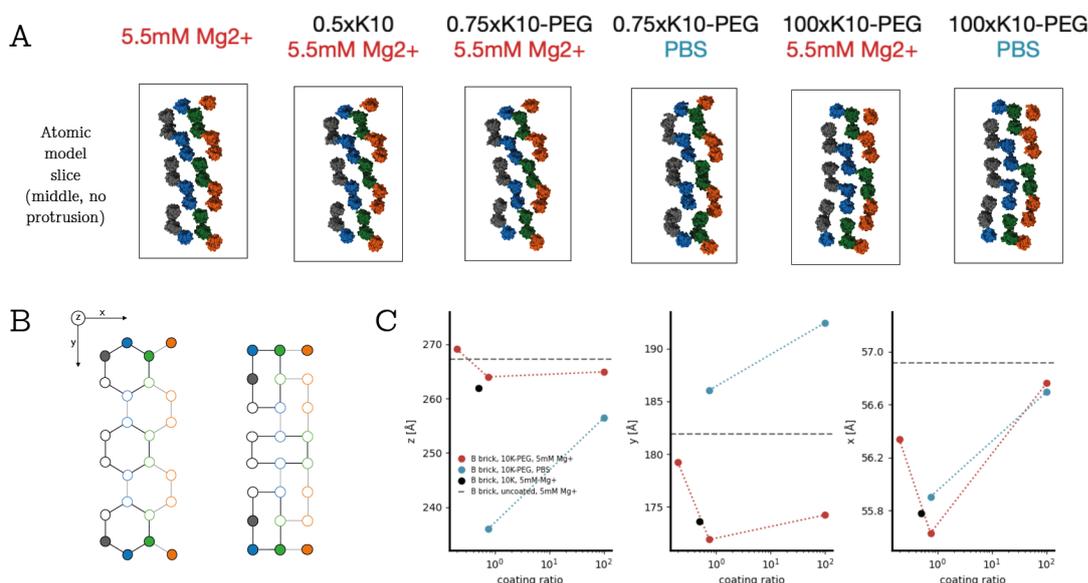
**Figure 4.2 | *B-Brick Structural Changes Due To Coating and Buffers Variation. (A)* *A 25 atomic model bases of the central slice (base position 85–100) at the coating and buffer conditions depicted in Figure 4.1. Helices of the pseudo-atomic model slices are colored by the honeycomb-lattice column (gray-blue-green-orange) of the caDNAno design file. (B) Structural transition of the helical pattern from regular honeycomb lattice (left) to compressed honeycomb (right). The compressed variant is obtained for the B-Brick structure coated with PEG-oligolysine at a 100:1 N/P ratio. (C) Structure dimension as a function of coating ratio. Directions x and y are calculated as an average of suitable helical pairs, while z is calculated as an average over all helical lengths. Data presented were calculated on the core unit (helices 6–29, base position 44–126), which excludes helical ends and protrusions from dimension measurements. Lines are guides for the eye. Adopted from [16]*

To quantify the changes in dimensions of the *B-Brick* (4.2C), we defined a core region that excludes the short helical protrusions of the dimerization motif and the loosely

connected helical ends from the analysis. Both omitted features are more flexible and often slightly deformed which reduces their overall resolution and would obfuscate the dimension measurements. We calculated helical distance in both lattice directions (x- and y-axis) and overall dimensions along the helices (z-axis). Measurements are averaged over the midpoints of the C1' atoms of each base-pair and overall helices, respectively.

All structures in a buffer with high ionic strength are more compact (x-axis and y-axis) than the uncoated structure, as the coating effectively shields the electrostatic repulsion of the backbone. Their length along the helical direction (z-axis) remains almost unchanged. This is even true for the thick coating with an N/P ratio of 100:1, although effects of the lattice transition are also observable. For structures in PBS, a length reduction of around 10% can be observed along the z-axis. A swelling of approximately 7% in the y-direction and no change in the x-direction is likely caused by the helical rearrangement of the lattice deformation.

4.2   Pseudo-Atomic Model-Guided Design

Remark: The work presented in this section is performed in collaboration with Khoshouei, A., Sachenbacher, K., and Kohler F. and contains cryo-EM data unpublished at the time of writing. Khoshouei, A. designed the research and performed cryo-EM measurements and data processing of the *Whopper* and *Whopper Junior* objects. Sachenbacher, K. developed the triplex mediated crossover motif and performed cryo-EM measurements of the *Triplex Bullet* objects. Kohler F. established the DNA origami designs of the 45° corners (*Corner Study Version 3*, *Corner Study Version 6*) and performed cryo-EM. The work presented on tubules assembly was published in Hayakawa et al. Geometrically programmed self-limited assembly of tubules using DNA origami colloids (2022) [170].

The availability of high-resolution volumetric electron microscopy data extends the iterative design pipeline for DNA origami nanostructures. Investigation of nano-objects on the nucleotide level enables optimization of individual helices. As duplex properties of DNA vary due to solvent interaction and local torques produced by the lattice [32], base-pair-level interventions by the designer are required to correct distortions of the global as well as local structure. The structural insights not only allow the validation of design goals but also high-precision engineering of additional motifs. These include additional crossovers to increase the integrity and rigidity of lattice-based origami, but also attachment of functional moieties for, among others, recognition of proteins [180], fluorescent signaling [11, 17], and the creation of artificial catalytic centers [181].

Here we present examples of research projects that used insights provided by pseudo-atomic modeling throughout the design process. The target objectives for utilization of topology and coordinates of the models can be assigned to three categories. Firstly, we compare different design versions using the positional data of the model. Secondly, we evaluate structural features to perform a high-precision design of additional motifs to extend an existing DNA origami. Lastly, we validate a predefined design goal by computing a quantifiable observable from the context of the quasi-atomic structure.

Comparison of cryo-EM data of different design iterations can be difficult because of the disparity in local resolution. Varying sample and ice quality, the micrograph orientation distribution, and the number of overall picked particles differ for each dataset. Consequently, intensity data for each voxel of the volumetric dataset fluctuate dependent on the relative position within the structure. At resolutions above 10 , this might even cause issues for precise alignment of two similar datasets using map-map cross-correlation. Additionally, the use of iso-surfaces for visual inspection and comparison of volumetric datasets is susceptible to misinterpretation. To overcome this issue, we can use pseudo-atomic coordinates for the alignment. Using the context provided by the model, we can quantify deviations of the models using root-mean-square deviations and reduce complexity by computing and comparing helical traces. Throughout this analysis, the linkage and *FitViewer* introduced in the previous Chapter 3 are used.

First, we investigate the influence of minor changes in the internal structure between the *Pointer* objects *Version 1* and *Version 2*. Both variants exhibit the same overall dimension, scaffold routing, and sequence. However, the designs differ in the set of staple oligonucleotides as Version2 uses predominantly half-crossovers instead of complete Holliday crossovers. The use of half-crossovers increases the number of possible staple-set combinations by increasing the number of strand termini placements. This enables a more homogeneous distribution of staple domains and overall staple length [85] and reduces the occurrence of backbone nicks. Using these design modifications, the yield and quality of the folding process could be increased for the *Pointer Version 2* [32]. To quantify the effect on the geometry of the DNA origami nanostructure, we built PAMs for both designs and used scaffold atoms to align them. To facilitate visual comparison we computed the helical traces (Figure 4.3A). The two versions do not show significant differences, emphasized by an overall scaffold RMSD of 7.49 Å, which is below cryo-EM data resolution. Half-crossovers, which have a nick instead of a backbone connection for one of the two strands, are also indistinguishable at this resolution, despite the anticipated reduction in rigidity. We conclude that the topology and sequence are the main factors for the structure
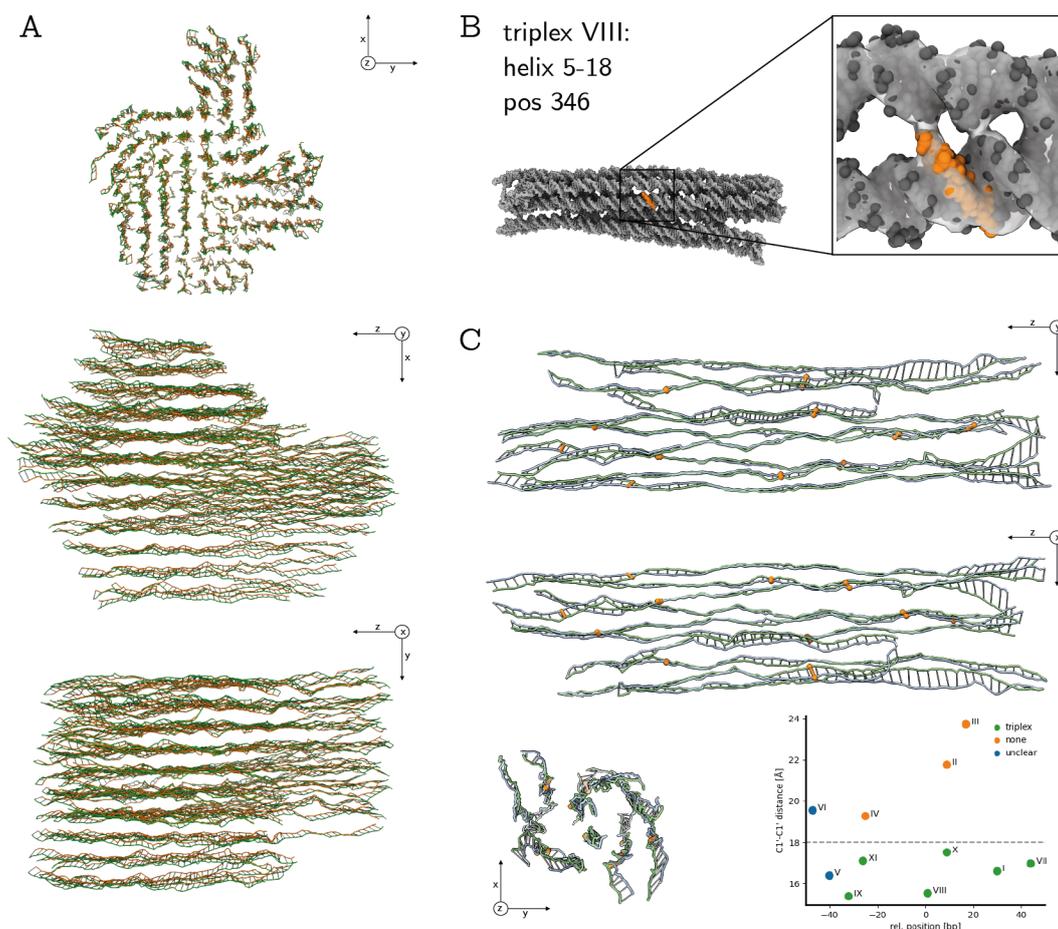
Figure 4.3 | Variant comparison. *(A) Helical trace comparison of the Pointer Version 1 (green) and Pointer Version 2 (orange). Lines connecting the trace lines (gray) depict the displacement at every second base position. (B) Overlay of cryo-EM map of the Triplex Bullet object with the quasi-atomic model of the duplex DNA. A manually added third strand (orange) indicates the triplex formation visible in the major groove. The rectangular box shows a magnification of a triplex-forming segment. (C) Comparison of helical trace lines of the Triplex Bullet PAM with (green) and without triplex formation (blue). Lines connecting the two models (orange) indicate designated Triplex binding positions. Inset: C1'-C1' distance of helical pair of the triplex-forming Triplex Bullet relative to the center of the design. Colors indicate the visibility of the triplex in the cryo-EM data. Distances were measured using coordinates provided by the PAM.*

of DNA origami nano-objects, while the type of crossover and distribution of backbone-nick sites is negligible. Importantly, this enables the optimization of the folding process using different strategies for staple placement and variation of the staple length distribution.

Second, we examine the effects of a novel design motif for a regular honeycomb-lattice DNA origami nanostructure. The *Triplex Bullet* object extends existing staples using a triplex-forming sequence that binds to the major groove of the adjacent helix [182, 183]

via Hoogsteen base-pairing [184]. For this design, eleven additional crossovers (I-XI) of these overhangs were placed at selected positions. Given the appropriate conditions, the staple extensions form triplexes to an adjacent helix, here referred to as triplex-crossovers. Using cryo-EM SPA we were able to validate the pH-dependent formation of triple-helical DNA. Figure 4.3B shows an exemplary triplex extracted using the *FitViewer* tool.

To quantify the potential structural change mediated by the formation of the triplexes, we build PAMs for two different volumetric datasets of the *Triplex Bullet*. The primary model was constructed using cryo-EM data of the original *Triplex Bullet* with helical triplets formed. The secondary model was built from the same PAM but using electron density data imaged from a variant of the design without the triplex staple overhangs. In both cases, triplex-forming oligonucleotides were excluded from the model, as the shrink-wrap framework of *dnaFit* dos not support Hoogsteen base-pairing. Nonetheless, the remaining double-helical segments of the origami lattice facilitate the assessment of the changes to the overall structure induced by the triplex formation.

The helical trace information computed from the PAM (compare Figure 4.3C) shows that the structure retains its overall shape. The RMSD between the two versions is with $9.02\,\text{Å}$ below the cryo-EM resolution. Comparison of the geometry of these models further facilitates quantification of the influence of the triplex formation. This is possible, as the two models are identical in terms of composition and topology. To this end, the distance between the helices at the position of the triplex-crossover was computed using the *dnaFit* Python framework. Using the distance measured between the C1' atoms (Figure 4.3C inset), we observe eleven data points with a C1'-C1' distance of less than $18\,\text{Å}$. All potential triplex locations below this threshold are classified as properly formed. This classification is also consistent with cryo-EM readout, with two triplex-crossovers not distinguishable experimentally due to the reduced resolution at the edges of the structure.

Third, we provide two examples for the utilization of pseudo-atomic models during the iterative design pipeline of high-precision DNA origami. For the first structure, we strive to reduce the relative motions of the two arms of the *Corner Study Version 6* object variants. Principle Component Analysis of the cryo-EM data reveals, that changes in the opening angle account for 23% of the total dataset variance. This relative motion [79] not only reduces the overall resolution but also impedes the accuracy of angle programming of the structure. To stabilize the angle at 45° double-stranded angle braces are to be placed between the corner arms to dampen the motion. These motifs are incorporated into the design by using reverse complementary staple sequences that extend from the inner helices of the two domains. To ensure the precision design of the angle the length of these duplexes and their placement is important. Braces of incorrect length either compress or
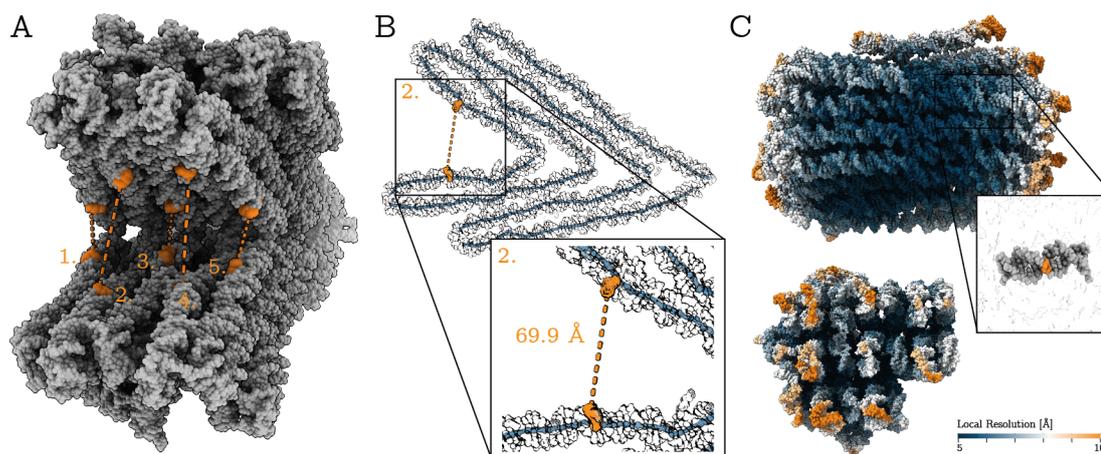
**Figure 4.4 | *Model-guided iterative design. (A)* Potential staple bases (orange) for the five angle braces of the *Corner Study Version 6* design (pseudo-atomic model, gray). *(B)* Exemplary layer slice of the second angle brace. At the highlighted staple base (orange), the two arms (PAM in gray, helical trace in blue) are 69.9 Å apart (zoom-in). *(C)* Whopper PAM colored by local resolution in helical (top) and cross-sectional view (bottom). The magnified subsection was extracted by the FitViewer as the 14-base segment with the best local resolution of all exterior helices. The segment belongs to helix 37 of the design diagram. The phosphate group of highlighted staple base (orange) at caDNAno position 73 points away from the DNA strand. It is a suitable candidate for moiety attachment.*

force the corner open, while a placement off the helical center (backbone position of the staple extension) can introduce torques that deform the overall shape. Here we used the generated PAMs to identify the optimal position for brace placement for the next design iteration. Additionally, distance measurement of the phosphate atoms of the placement sides on each corner arm facilitates specification of the number of base-pairs required for each angle brace (compare Figure 4.4A-B).

The second design goal requires the identification of the ideal position for the placement of a protein binding moiety. Analysis of the *Whopper Junior* object is shown exemplarily for the group of functional RNA platform objects. Like in the previous example, the context provided by the atomic model enables high precision selection of the appropriate base position. Using the *dnaFit* package, we can generate atomic structure files (.pdb) that contain an additional property of the model describing the experimental data. In this case, we encoded the local resolution information provided by cryo-EM processing with RELION [97, 98] into the occupancy field of the structure file. This enables color-coding of the pseudo-atomic model by local resolution (Figure 4.4C). Using this approach, we identified a suitable position on one of the exterior helices with maximum local resolution and staple backbone pointing outwards.
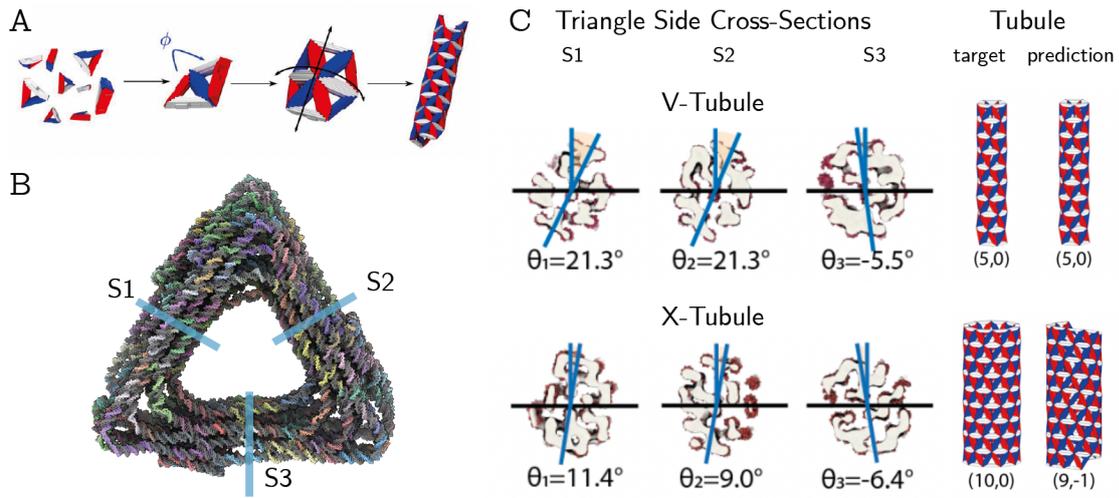
Figure 4.5 | *Tubule forming triangle validation. (A)* ** *Tubule formation by the assembly of triangular subunits. The binding angle ϕ of the triangle specifies the curvature of the cylindrical tubules. **(B) PAM of the V-Triangle fitted into electron density using the low-resolution shrink-wrap protocol. The markers (blue) indicate the three distinct triangle sides. **(C)** Measurement of the dihedral angle of each of the three sides of the V-Triangle (left, top) and X-Triangle (left, bottom) was performed using atomic coordinates of the PAM. The depicted tubules illustrate target geometry (right, target) and the expected tubule geometry (right, prediction) based on the model measurements. Panel (A) and (C) adopted and modified from [170].*

Finally, Hayakawa et al. [170] use DNA origami to design triangular subunits for tubule assembly (compare Figure 4.5A). The diameter and the chirality of the resulting tubules can be geometrically programmed by tuning the dihedral angles between neighboring subunits. Cylindrical tubules resulting from this procedure can be characterized by a tuple (m, n), with m denoting the number of subunits along the direction of the circumference and n the offset along the second direction of the planar lattice described by an unraveled tubule. Negative n results in left-handedness, positive n in right-handedness, whereas n=0 defines an achiral tubule. Pseudo-atomic models of two different triangular subunits were used to quantify and validate the designed dihedral angle (compare Figure 4.5B). Due to an SPA resolution of >15 Å, the low-resolution protocol described in Section 2.2 was utilized. Atomic coordinates generated by the model building procedure were then used to define and compute the three side planes and their angles $(\Theta_1, \Theta_2, \Theta_3)$ relative to the orientation of the unit. In order to mitigate the increased flexibility of helices on the exterior of the structure, planes were defined on the interior column of helices for each of the three sides. Protrusions that form the lock and key mechanism for programmable connection of the subunits were also removed.

An illustration of the dihedral angle determination and a comparison between designed and experimentally determined target tubules are shown in Figure 4.5C. The *V-Triangle*

structure was designed to form an achiral (5,0) tubule, which requires the target angles
$(20.9, 20.9, -10.8)$. The *X-Triangle* object is also achiral but exhibits a larger diameter
with (10,0) tubule, resulting in a set of angles half as big $(10.4, 10.4, -5.3)$. The angles
computed from the SPA cryo-EM data of the monomer of the *V-Triangle* are in agreement
with the designed values at $(21.3 \pm 0.1, 21.3 \pm 0.1, -5.5 \pm 0.1)$. However, angles for the
X-Triangle at $(11.4 \pm 0.1, 9.0 \pm 0.1, -6.4 \pm 0.1)$ are closer to the set forming a (9,-1) tubule.
While (9,-1) and (10,0) tubules are comparable in overall diameter, the experimentally
determined values yield a left-handed chiral tubule.

Psuedo-atomic models of the *A-Triangle* and *B-Triangle*, presented in Section 2.3, are
designed to form a binary species (10,0) tubule [170]. However, experimental data showed a
significant deviation from the target angles of the *A-Triangle*, likely caused by the complex
lock and key mechanism required for dual-species tubules.

4.3 B-DNA Duplex Properties

Within the biological context, the sequence-dependent base-pair level deformability of DNA
is an essential mechanism for the binding of proteins and regulation of gene expression.
Similarly, the effects of its sequence on the conformation of a DNA duplex gain importance
as the method of DNA origami progresses towards higher precision design. Our atomic-level
models, constructed into cryo-EM data, provide a large set of base coordinates of DNA
duplexes in solution. Based on these coordinates we compute the base- and base-pair
properties of B-DNA for a set of almost 40 thousand base-pair steps.

The analysis pipeline of *dnaFit* implements an interface (see Section A.3 for details) for the
computation of duplex properties based on a standard reference frame [64, 65]. Sequence-
dependent values for the six base-pair step properties are listed in Table 4.1. The dataset
presented in this section was generated from the *B-Brick*, both *Pointer* versions, both
Twisttower variants, the 48Helix-Bundle, and the *126-Helix Bundle* objects. Steps involving
crossover motifs were excluded from the analysis. Base-pair step properties were computed
by the mean and standard deviation of a harmonic approximation of their interaction [185]
using a Gaussian function.

In comparison to pure *enrgMD* structure prediction duplex properties (see Appendix B.3)
only minimal variation of the mean values was observed. In general, we observe a stronger
sequence-dependent variability of the mean, than permitted by the elastic network of
enrgMD. Species distributions are broader than compared to the structure prediction.
Noteworthy, base-pair enforcing bonds of the elastic network are still present even at the
final stage of the shrink-wrap protocol. We are encouraged by the deviation from *enrgMD*

Step	Twist [°]	Roll [°]	Tilt [°]	Rise [Å]	Shift [Å]	Slide [Å]
AA	38.05 ±4.16	37.99 ±4.68	13.95 ±8.11	3.15 ±0.72	-0.90 ±0.92	-0.12 ±0.58
AC	35.33 ±6.49	37.67 ±6.26	14.86 ±9.07	3.08 ±0.73	-1.36 ±0.75	0.04 ±0.77
AG	37.69 ±5.27	37.75 ±6.01	11.77 ±7.08	3.40 ±0.59	-0.79 ±0.89	-0.12 ±0.65
AT	34.99 ±6.21	36.54 ±6.76	14.40 ±8.93	3.37 ±0.71	-0.80 ±0.96	-0.35 ±0.70
CA	35.41 ±6.38	36.75 ±6.36	12.81 ±8.24	3.57 ±0.74	-0.38 ±1.15	-0.18 ±0.81
GA	37.23 ±5.65	38.05 ±5.59	12.01 ±7.79	3.36 ±0.74	-0.62 ±0.96	0.05 ±0.71
GC	35.56 ±6.70	37.73 ±6.36	12.67 ±8.81	3.10 ±0.82	-0.99 ±0.92	-0.31 ±0.69
GG	38.76 ±5.90	39.98 ±6.90	13.14 ±8.61	3.40 ±0.76	-0.52 ±1.44	0.10 ±0.94
–	36.85 ±5.76	37.77 ±5.96	13.27 ±8.35	3.30 ±0.74	-0.82 ±1.01	-0.12 ±0.72

Table 4.1 | _Base-pair step properties._ _Sequence-dependent base-pair step parameters (three rotational, three translational) and overall average. Average values are computed using a uniform base-step distribution. Reported error values represent the spread calculated assuming Gaussian distributions._

of sequence-depended properties, despite not being able to reproduce trends reported in previous studies [185–188]. Additionally, we were not able to replicate multi-modal variation of tetranucleotide sequences [70, 189] caused by backbone structure polymorphism [190–192]. As individual base-pairs are not yet distinguishable visually at cryo-EM resolutions around 10 Å, we anticipate that higher resolution will be required before the data accurately reflect these effects. Further improvement in SPA of DNA origami nanostructures will likewise enable the removal of the remaining elastic network restraints present during the model building, removing its bias.

4.4   Conclusion

Throughout this chapter, we have presented a collection of independent projects that utilized shrink-wrap fitting for structural analysis and iterative design of DNA nanostructures. With the goal of each project different, these examples highlight the versatility of the annotation and context provided by the pseudo-atomic model.

In summary, we have used the geometric data of the PAMs to quantify overall dimensions and structural changes caused by the coating of DNA structures with cationic cofactors and variation of buffer conditions. We compared two variants of the _Pointer_ design to determine that the overall dimensions of compact lattice-based DNA structure are not strongly affected by the configuration of the staple set as long as the sequence and topology are unchanged. Similarly, the same atomic models fitted into two different cryo-EM datasets of the _Triplex Bullet_ facilitated the measurement of structural changes mediated by the addition of triplex-forming staple extension into a regular honeycomb-lattice design.

Moreover, we demonstrated how positional information provided by the model-to-design linkage can be utilized for precision placement of moieties or structural elements during the iterative design process of the *Whopper* and *Corner Study Version 6*, respectively. The computation of dihedral angles for the accurate prediction of DNA origami *B-Triangle* ased tubule assembly highlights the potential of PAM building to answer target-specific research questions. Finally, we presented the utility of *dnaFit* to compute sequence-dependent duplex properties of B-DNA, by computing base-pair step properties and comparing them to elastic network-driven all-atom structure prediction results. Although the overall cryo-EM resolution is not yet sufficient to replicate the multimodality of tetranucleotide sequences, data obtained from a set of almost 40 thousand base-pair steps show variation from the prediction reference.

The list of collaboration partners at the beginning of each section accentuates the versatility and demand for structural annotation and exploratory data analysis of cryo-EM volumetric data. The presented cases display the importance of interdisciplinary synergy for modern biophysics in general and the field of DNA nanotechnology in particular.

Chapter 5

Structure Prediction Validation

Equilibrium ensemble structure prediction of DNA origami nano-objects has been limited for years by the small number of available atomic coordinates. For almost a full decade, model parametrization relied on a single pseudo-atomic model [33], low resolution Atomic Force Microscopy data [3, 27] and MD simulation force fields based on experimental Nuclear Magnetic Resonance data of DNA duplexes [54–56, 75, 193]. Aside from the elastic network-guided atomic resolution model *enrgMD* [31], utilized for shrink-wrap fitting in Chapter 2, several coarse-grained models suitable for DNA origami have been developed. Among them are the Finite Element models CanDo [26, 126, 127] and its improved version *SNUPI* [39], as well as the coarse MD force fields *oxDNA* [38, 118], *mrDNA* [116], and COSM [167]. The quasi-atomic model of the *Pointer Version 1* object served as the main resource for comparing high-resolution features of the simulation results of these models. Exceptions are CanDo, which predates the *Bai et al.*, and *oxDNA*, which was not originally developed for structure prediction of DNA origami [119, 120], but later improved and extensively tested for the field of DNA nanotechnology [76, 194]. Validation of these novel prediction models was hence limited, and it remained unclear how well they generalize for arbitrary DNA origami designs.

With shrink-wrap fitting, we used the cryo-EM data of the *Pointer Version 1* to validate the correctness of our protocol. The semi-automatic fitting procedure expands the high-resolution geometric data available for validation of structure prediction methods. So far, we published eight pseudo-atomic structures to the Protein Data Bank [158] (identifiers are listed in the Appendix Table B.1), with several more currently in the process of publication. These designs now cover a broad spectrum of overall size, cross-section, lattice type, design motifs, and global twist. Consequently, the topological and geometric data of these models form a suitable dataset for the validation of prediction models and their parametrization.

In this chapter, we present a comprehensive overview of the predictive power of the three categories of aforementioned structure prediction models: atomic resolution MD, coarse-grained MD, and coarse-grained Finite Element simulation. The main focus of this study lies with the *return of investment* of each prediction type, signifying the predictive accuracy in comparison to the compute time spent on the simulations. First, we compare our pseudo-atomic models (PAMs) with the initial structure prediction of *enrgMD*. Then, we simulated a selected set of designs with the most commonly used [195, 196] coarse-grained structure predictions methods of the remaining two categories, *oxDNA* and *SNUPI*. Thereby, we establish a set of reference predictions intended to serve as a guideline for determining the appropriate prediction method for various workflows in the field of DNA origami nanotechnology.

5.1 All-Atom Comparison

Remark: The work presented in this chapter was published as supplementary information in Feigl, Kube, Kohler, et al. Revealing the structures of megadalton-scale DNA complexes with nucleotide resolution (2020) [32].

Even for small DNA constructs the computational cost associated with explicit solvation presents an obstacle for simulation studies of these systems [197, 198]. For *enrgMD*, the replacement of the solvent with an elastic network not only removes a majority of interaction sites from the system but also allows for faster dynamics using a 2 fs time step. With its development, the computational cost of atomic level simulation could be reduced by five orders of magnitude [31]. This drastic decrease makes *enrgMD* atomic-level structure prediction a viable tool for the DNA origami design process [13, 86, 199]. The speed-up for overall dynamics and domain motions induced by the elastic network is one of the main reasons for including *enrgMD* as a central part of the fitting procedure outlined in Section 2.1.

As mentioned previously, the *Pointer Version 1* object served as the primary source for the parametrization and validation of the elastic network. To assess how well the model generalizes we compared all-atom *enrgMD* structure predictions of six different DNA origami designs with their quasi-atomic models. Overall resolution and RMSD of prediction and cryo-EM derived model are listed in table 5.1. In general, all values are above the resolution of the experimental maps used for fitting. For three of the tested models, RMSD values are below the reported resolution of 12 Å for the *Pointer Version 1* object [33], indicating excellent predictive power. As the elastic network was parametrized based

[Å]	Resolution	RMSD
Pointer Version 2	7.4	9.66
Twisttower Native Twist	7.0	16.46
Twisttower Corrected Twist	8.1	10.99
48-Helix Brick	9.9	14.93
16-Helix Bundle	10.0	10.31
126-Helix Bundle	8.8	13.56

Table 5.1 | *Deviation of enrgMD atomic level prediction and cryo-EM derived model.* *Overall resolution compared to all-atom RMSD of a subset of six models originally published in [32]. Resolution estimates are computed via Fourier Shell Correlation using RELION [97, 98]. RMSD values were obtained in VMD [166] from all atoms after alignment.*

on data of this resolution, it represents the expected level of accuracy for the prediction method. With a deviation of less than 9.7 Å, the prediction for *Pointer Version 2* matches experimental data closest, which is to be expected given its similarities to *Pointer Version 1*. Nonetheless, it serves as an indicator of the independence of the structure of the model from staple breaking strategies, as long as the overall connectivity remains unchanged. Structure prediction with the elastic network also reliably predicts honeycomb-lattice structures with RMSD < 15 Å, with the *48-Helix Brick* object generating the largest deviation. This highlights the versatility of the elastic network since at the time of its development no solution structure for honeycomb-lattice was available.

The strongest deviation for any model tested occurred for the *Twisttower Native Twist*, which coincidentally reports the best overall resolution at 7 Å of the model subset. To further assess the cause of these deviations, we compared the helical traces of the atomic coordinates. For the *Twisttower Native Twist* object (Figure 5.1A) strong global right-handed twist deformation and buckling of the individual domains of the different helical cross-sections are the main cause of the deviation. As the *Pointer Version 1* object (Figure 5.1D) does not exhibit any significant twist, the elastic network appears to not adequately reflect this property. In addition, the twist deformation causes a buckling of the domains at sites with strong changes in cross-section, which in turn impedes proper global alignment of the structure. The effect is most notably in the 2×2 domain with its strong twist and worst domain alignment. However, the twist-corrected variant of the *Twisttower* (Figure 5.1B) is among the best predictions of the set since it does not deviate as drastically from the idealized lattice grid.

Strong deviations between the fitted model and atomic-level prediction also occur at positions of omitted crossovers, especially for exterior helices. Lacking topological connectivity, these helices tend to bend away from the main body due to the electrostatic repulsion of the DNA backbone. In *enrgMD* long-range interactions are replaced by harmonic
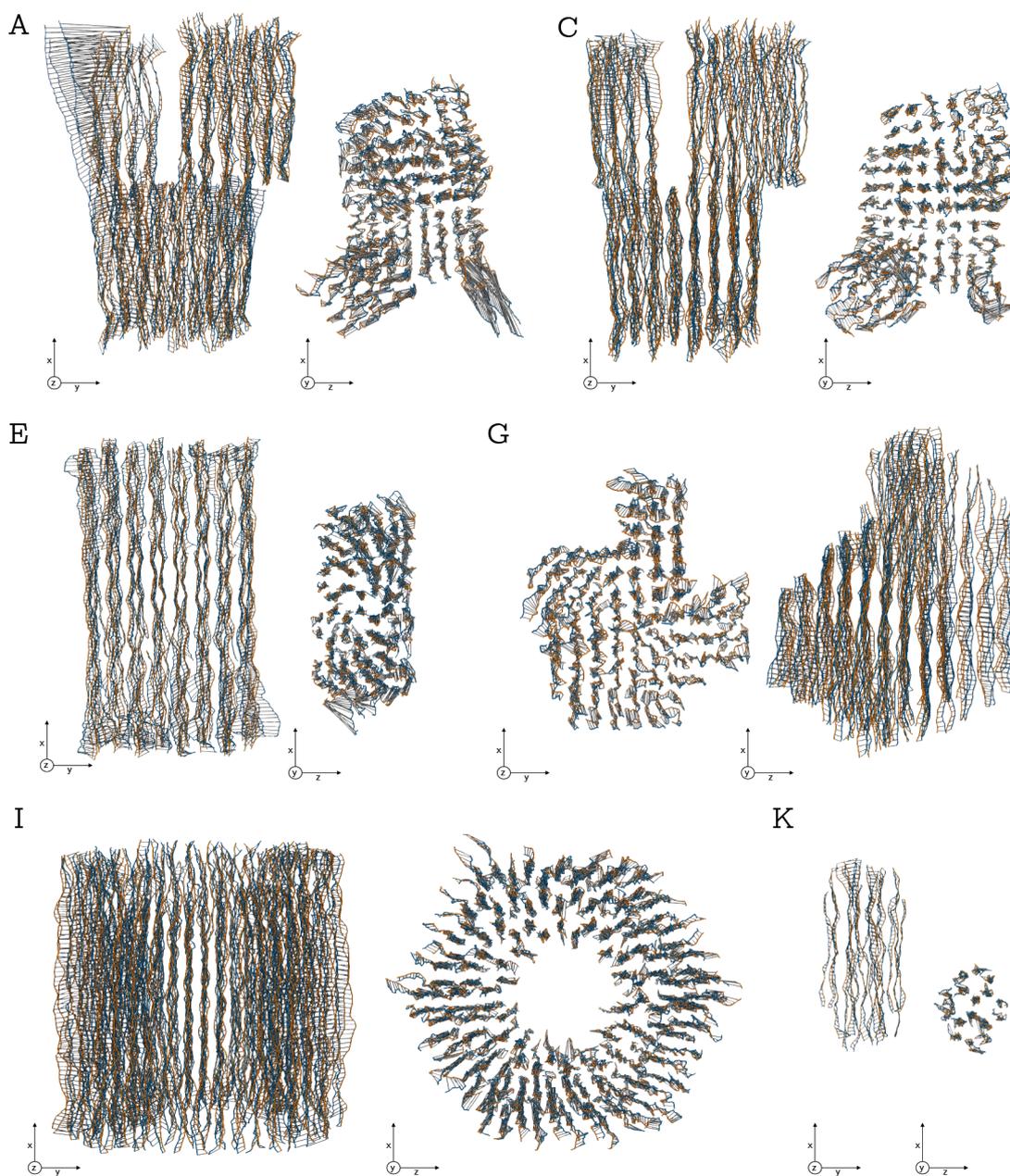
Figure 5.1 | *Helical trace comparison between enrgMD and shrink-wrap-fitting.* [32]
Globally aligned helical trace lines of the pseudo-atomic models (orange) and enrgMD structure
prediction (blue) for the models: (A) Twisttower Native Twist, (B) Twisttower Corrected Twist,
(C) 48-Helix Brick, (D) Pointer Version 2, (E) 126-Helix Bundle and (F) 16-Helix Bundle.
Displacement at every second base position is depicted with connecting lines (gray).

potentials of the elastic network. Consequently, the same helices are either tied to the
structure by the restraints of the elastic network or not repulsed strong enough due to the
lack of long-range interaction, even if no restraint is present. The *48-Helix Brick* design

(Figure 5.1C) shows this flaring feature for two of its corner helices, causing larger RMSD than the two remaining honeycomb-lattice structures (Figure 5.1E-F).

5.2 Coarse Grained Comparison

For most use-cases in DNA origami design, atomic level predictions are not necessary. Moreover, the high computational cost, long simulation times, and expertise required for all-atom MD might make its use impractical. At the start of the development of novel nano-objects, properties like the overall shape and dimensions necessitate verification. The analysis of early iterations of a design depends on the topology of the macromolecules and sufficient connectivity of the individual helices. As the design progresses, base-pair level details are often the focus of improvement and validation of the design.

Exemplarily, the flexibility of a subdomain or the placement of an additional design motif might not be apparent from the topology diagram alone. Reviewing the structures presented in the previous chapters, we compiled a list of hypothetical questions that might have presented themselves during their design. All these questions can be answered using fast, coarse-grained structure prediction and do not necessitate atomic-level resolution.

- *Twisttower Corrected Twist*: Are all four domains completely free of a global right-handed twist? Has its correction induced any local deformations of the lattice?
- *16-Helix Bundle*: Given its small size, are all helical ends sufficiently connected to form a compact and rigid structure?
- *Corner Study Version 3*: Do the two arms of the corner comprise a 45° angle?
- *126-Helix Bundle*: Are crossovers formed by two scaffold strands distinguishable from regular scaffold crossovers? How big is the diameter of the hexagonal inner cavity o the dual-scaffold structure?

In this section, we present the prediction results of a set of ten reference structures. All of these designs have been imaged using cryo-EM and a PAM is available. This enables the comparison of the prediction results with the experimental data of the model, quantifying their predictive power.

Structure predictions were performed with both *oxDNA* and *SNUPI*. *oxDNA* provides a high level of accuracy within a fraction of the compute time required for all-atom simulations like *enrgMD*. As a coarse-grained MD simulation, developed independently of DNA origami lattices, it is highly flexible and versatile. All *oxDNA* simulations were performed with model version 2 which includes, among others, salt dependence and helical grooving [76].

Additionally, the compatibility of *oxDNA* with the multi-resolution *mrDNA* [116] facilitates a speedup and automatization of the relaxation procedure.

In contrast, the finite elements of *SNUPI* are optimized or wireframe [6] and lattice-based origami. Utilization of the Finite Element Method makes *SNUPI* significantly faster than MD-based prediction models. While slightly slower than its predecessor CanDo [26], the added complexity introduced by *SNUPI* makes it a more viable rapid prediction tool. For *SNUPI* predictions version 2.1 was used, which features the partition & relocation [128] framework. Both *SNUPI* and *oxDNA* are actively developed, constantly improved, and updated.

5.2.1   Validation Structure Set and Protocol

To represent an adequate cross-section of the variation in DNA origami nanostructure designs, we selected a set of ten structures: *Pointer Version 2*, *Twisttower Native Twist*, *Twisttower Corrected Twist*, *Whopper*, *Whopper Junior*, *16-Helix Bundle*, *48-Helix Brick*, *126-Helix Bundle*, *Corner Study Version 3*, and *B-Triangle*. Three of these objects are built in honeycomb-, seven in square-lattice. Their overall size ranges from a 1033 base scaffold of the *Whopper Junior* to scaffolds of 8064 bases of the two *Twisttower* variants, *48-Helix Brick*, and *B-Triangle*. The dual-scaffold object *126-Helix Bundle* is even bigger, with two times 7560 scaffold bases. The *Twisttower* variants represent a test set for the overall twist of square-lattice structures, while the *Corner Study Version 3* and *B-Triangle* designs represent open and closed angled objects. Designs composed of angled subdomains are included to test the relaxation procedures of *mrDNA* and the partition & relocation framework of *SNUPI*. Additionally, some objects contain special design motifs, for example, the large crossover stacks of the *Twisttower Native Twist* or the dual-scaffold crossover of the *126-Helix Bundle*. The *Pointer* design, which has been used for parametrization of both prediction models is included as a baseline reference.

All prediction models are compared to the PAMs generated with the regular shrink-wrap fitting protocol, except the *B-Triangle* for which the model was built using the low-resolution mode. As the abstraction levels of the tools are different, a direct comparison of the results is not possible. To facilitate numerical comparison, atomic structure (.pdb) files are generated as the final output for the predictions. However, these atomic coordinate files have two distinct drawbacks. First, atomic coordinates generated from both *SNUPI* and *oxDNA* do not reflect simulated coordinates but are interpolated rigid nucleotide models that are placed according to *oxDNA* base or finite element positions. Consequently, these configurations can include steric clashes and nonphysical bond lengths. Secondly, each

program uses its unique pdb-generation script, producing disparate quasi-atomic models. These models not only differ in atom and strand numbering but also in nomenclature and even the total number of atoms in the system (for instance, *SNUPI* places a phosphate at both 5' and 3' ends of a staple, while tac*oxDNA* does not).

To counteract these drawbacks, we used a modified version of *dnaFit* to compute helical trace data (see Appendix B.2.1) for all base-pairs of a structure. These helical midpoints represent the duplex on a level of complexity that is compatible with the two coarse-grained structure prediction techniques and the atom-level representation of *enrgMD* and the model building protocol. Reducing the complexity for comparison also helps mitigate effects on the RMSD stemming from the PDB generation. We used the helical trace coordinates to align the structures and compute their RMSD. These RMSD values, combined with the compute time for each structure and tool form the core metrics for assessment of the predictive power of *oxDNA* and *SNUPI*. Helical trace data are also used for visual comparison of the overall structure and design motifs. Simulation protocols and parameter sets are documented in Section A.4. Importantly, *SNUPI* simulations were performed with an electrostatic cutoff twice as large as the default value, resulting in a performance penalty factor of up to four.

5.2.2   Predictive Power

Simulations for each of the ten DNA origami designs were completed successfully. Timing data for the coarse-grained simulations are reported in Table B.4 of the Appendix. For structure prediction with *SNUPI*, the simulation times range from half a minute for the smallest structures to approximately 15 minutes for structures with the 8064 base scaffold. Two structures exceeded this time frame with the *Twisttower Native Twist* at around 26 minutes and the *126-Helix Bundle* at just over 2.5 hours. In the case of the *Twisttower*, the increased duration is likely caused by the large crossover stacks that slow down the convergence rate. The long simulation time for the *126-Helix Bundle* is rooted in the non-linear increase of computation time with the size of the system. Similarly, the dual-scaffold structure took around 84 hours for the *mrDNA/oxDNA* protocol, significantly longer than the 10 to 31 hours for regular designs. Overall, *SNUPI* is about two orders of magnitude faster.

The RMSD values between structure prediction and PAM are reported in Table 5.2. All *oxDNA* simulations have a trace RMSD better than 16 Å, with an average of 11.1 Å. *SNUPI* performs almost as well with all designs better than 19 Å and an average of 14.4 Å. In both cases, the best results are achieved for the *Pointer Version 2* design. Doubling of the default electrostatics cutoff with *SNUPI* increased the prediction accuracy by 1.2 Å on

[Å]	oxDNA	SNUPI
Pointer Version 2	6.89	7.91
Twisttower Native Twist	14.02	11.02
Twisttower Corrected Twist	11.99	12.48
Whopper	10.03	10.06
Whopper Junior	8.04	9.35
16-Helix Bundle	9.37	11.14
48-Helix Brick	15.43	18.94
126-Helix Bundle	11.25	11.25
Corner Study Version 3	10.71	15.51
B-Triangle	13.23	15.92

Table 5.2 | Trace RMSD: Coarse Grained Models. *Helical trace RMSD for the mrDNA/oxDNA prediction protocol and SNUPI2.1 with an increased electrostatic cutoff of a subset of ten DNA origami models. Trace data of the pseudo-atomic models of each model were used as reference.*

average compared to the default settings (see Appendix B.3) with significant improvements for the *Twisttower Native Twist* (22.7 to 11.0 Å) and *126-Helix Bundle* (19.9 to 11.3 Å).

5.2.3 Honeycomb and Square Lattice

Following the overall comparison of RMSD values, we investigated the dependence of the predictive power on the lattice type of the DNA origami nanostructure. For the honeycomb-lattice both the small *16-Helix Bundle* and the dual-scaffold, *126-Helix Bundle* reach RMSD values below 12 Å. The cavity diameter of the 126-Helix Bundle (Figure 5.2C) deviates by 10 Å for *SNUPI* and 6 Å for *oxDNA* from the value of approximately 171 Å measured on the PAM. For the *16-Helix Bundle* simulations with both tools predict the flaring of helical ends (see Appendix Figure B.4) also observed in the cryo-EM data. The *48-Helix Brick* is predicted well at RMSD values below 19 Å, yet cannot reach the same accuracy as the other objects. A comparison of the cross-section of the helical trace data at the two ends of the structure (Figure 5.2B) reveals that both *SNUPI* and *oxDNA* do not represent the slight right-handed twist of honeycomb structures with rectangular cross-sections [32]. Global misalignment combined with the flaring edge-helices (compare *48-Helix Brick* Section 5.1) result in a strong deviation. The RMSD of the *126-Helix Bundle* is close to the cryo-EM resolution of 9.8 Å [32] despite the size of the design. Both prediction models are capable of handling designs with multiple scaffold strands.
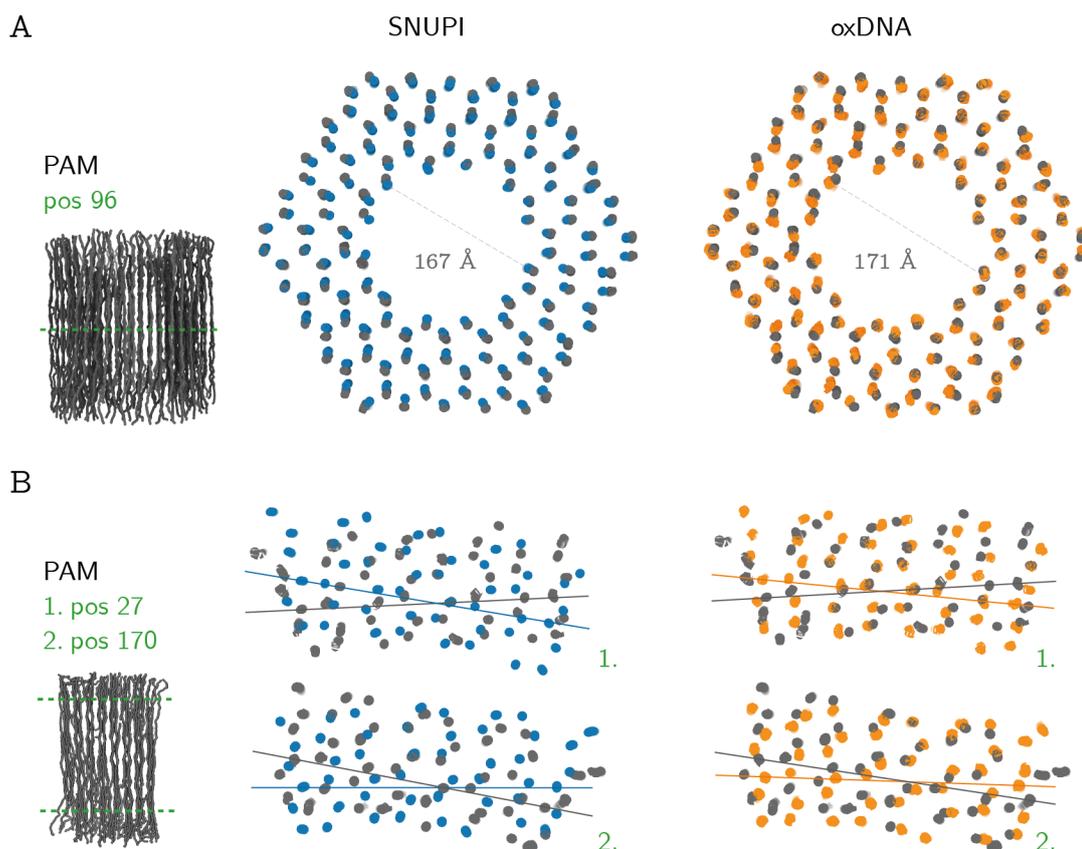
Figure 5.2 | *Coarse prediction for honeycomb-lattice.* *Visual comparison of the cross-sectional slice of the reference PAM (gray) to SNUPI (blue, middle) and oxDNA (orange, right).* ***(A)*** *126-Helix Bundle object sliced in the center around position 96 of the caDNAno diagram.* ***(B)*** *48-Helix Brick design with two slices at each end around positions 27 and 170 of the caDNAno diagram. Lines drawn from helix 3 to helix 44 of the design diagram indicate a global twist and are colored according to the model. Slice markers (green, dashed) indicate the position of the cross-section relative to the PAM (left).*

The globular square-lattice nano-objects *Whopper Junior*, *Whopper*, and *Pointer Version 2* did yield the best RMSD values overall with both *oxDNA* and *SNUPI* at 10 Å or better (see Figure 5.3).

In general, the lattice type did not show a significant effect on the predictive power, with only minor deviations for the potential global right-handed twist deformation of honeycomb-lattice. The larger cut-off for the electrostatic interaction of *SNUPI* appears to benefit larger structures more than structures with less than 3000 scaffold bases. The effect was particularly noticeable for the *126-Helix Bundle* with an RMSD improvement of 8.6 Å. This strong discrepancy is likely due to the large diameter and the barrel-shaped cross-section.
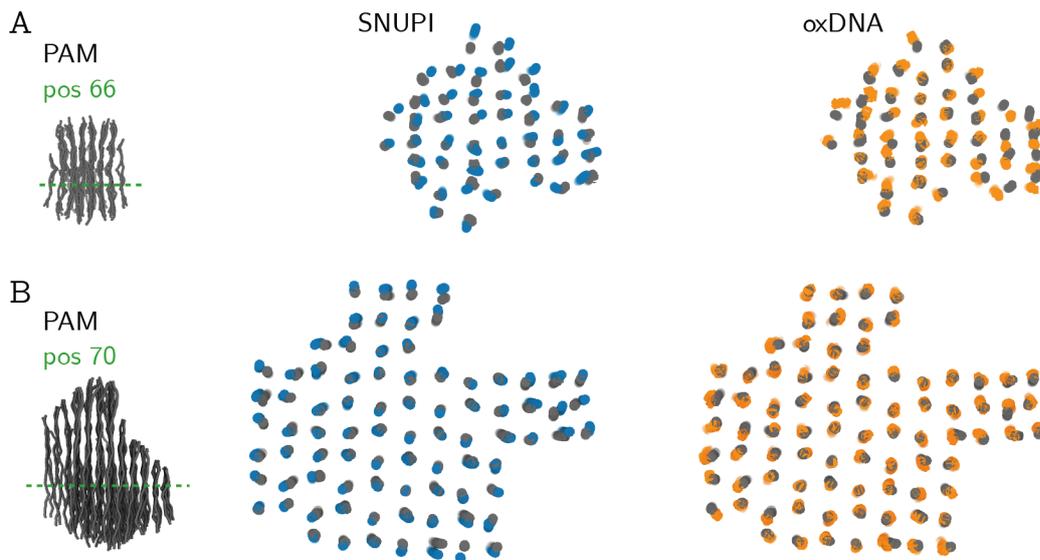
**Figure 5.3** │ ***Coarse prediction for square-lattice.*** *Visual comparison of the cross-sectional slice of the reference PAM (gray) to SNUPI (blue, middle) and oxDNA (orange, right).* **(A)** *Whopper design sliced in the center around position 66 of the caDNAno diagram.* **(B)** *Pointer Version 2 design sliced in the center around position 70 of the caDNAno diagram. Slice markers (green, dashed) indicate the position of the cross-section relative to the PAM (left).*

5.2.4   Overall Twist and Non-Standard Design Motifs

Deviation values around $12\,\text{Å}$ for both *Twisttower* variants suggest adequate prediction of the twist-deformation of square-lattice [32]. Analysis of the cross-section at both ends of each domain of the *Twisttower Native Twist* (Figure 5.4A) confirms a correct twist prediction for all but the 2×2 domain for *SNUPI* and only minor deviations in the 4×4 domain for *oxDNA*. Inaccurate prediction of the 2×2 domain with *SNUPI* is to be expected due to the low crossover density of this subdomain which causes disparity from the regular lattice. The alignment of the individual subdomains is in both cases limited by the electrostatic long-range interaction. Alignment of the 2×2 domain is better for *SNUPI* than for *oxDNA*, with *SNUPI* simulations performed at a larger electrostatic cutoff of $4\lambda_D$ compared to *oxDNA* at $3\lambda_D$ [76].

In case of the *Twisttower Corrected Twist* version (Figure 5.4B), both models show comparable results. The correction of the overall twist deformation is accurately represented for both approaches. Helical spacing appears slightly more accurate for *oxDNA*, resulting the lower RMSD values.
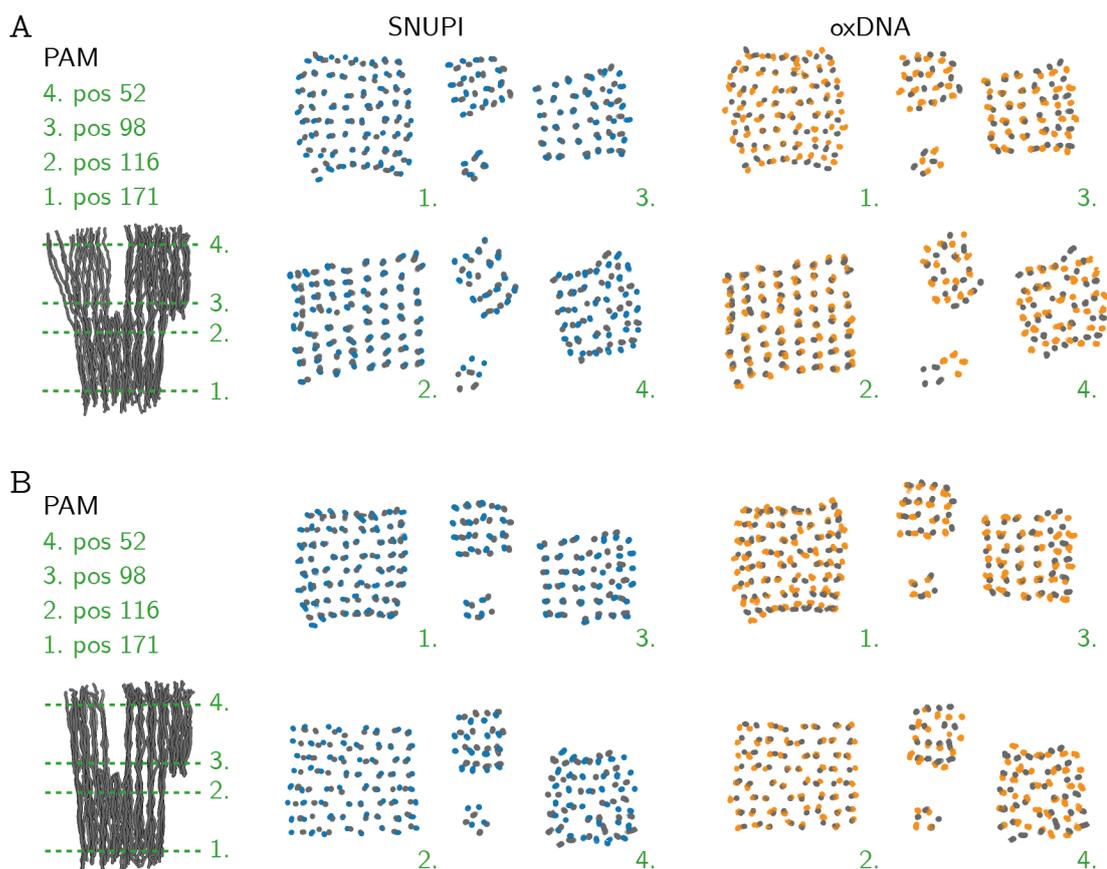
Figure 5.4 | *Coarse global twist prediction.* *Comparison of the cross-sectional slices of both variants of the Twisttower object. Reference PAM (gray), SNUPI (blue, middle), and oxDNA (orange, right) predictions are displayed at four positions that mark both ends of the four subdomains.* **(A)** *Twisttower Native Twist at both ends of the 8 × 8 domain (caDNAno position 171 & 116) and small domains (caDNAno position 98 & 52).* **(B)** *Twisttower Corrected Twist similar to (A), caDNAno positions are identical. Slice markers (green, dashed) and numbering indicate the position of the cross-section relative to the PAM (left).*

Additionally, both *oxDNA* and *SNUPI* predict the barrel-like distortion [32, 79] of the lower part of the 8 × 8 domain. These distortions are caused by crossover stacks that span the complete cross-section. No covalent bond of a DNA strand is connecting helical segments of such a crossover-stack site and the nanostructure is connected by blunt-end $\pi - \pi$ orbital stacking only. Since such stacking interactions are not implemented in the set of finite elements of *SNUPI*, these parts of the structure cleave, resulting in a prediction artifact (Appendix Figure B.5A). The missing interaction between the helices can also impact the convergence times of the prediction.

The dual-scaffold crossover motif of the *126-Helix Bundle* [91] does not differ from a regular scaffold crossover for both coarse-grained models. This result is consistent with high-resolution cryo-EM data and the PAM of the design.
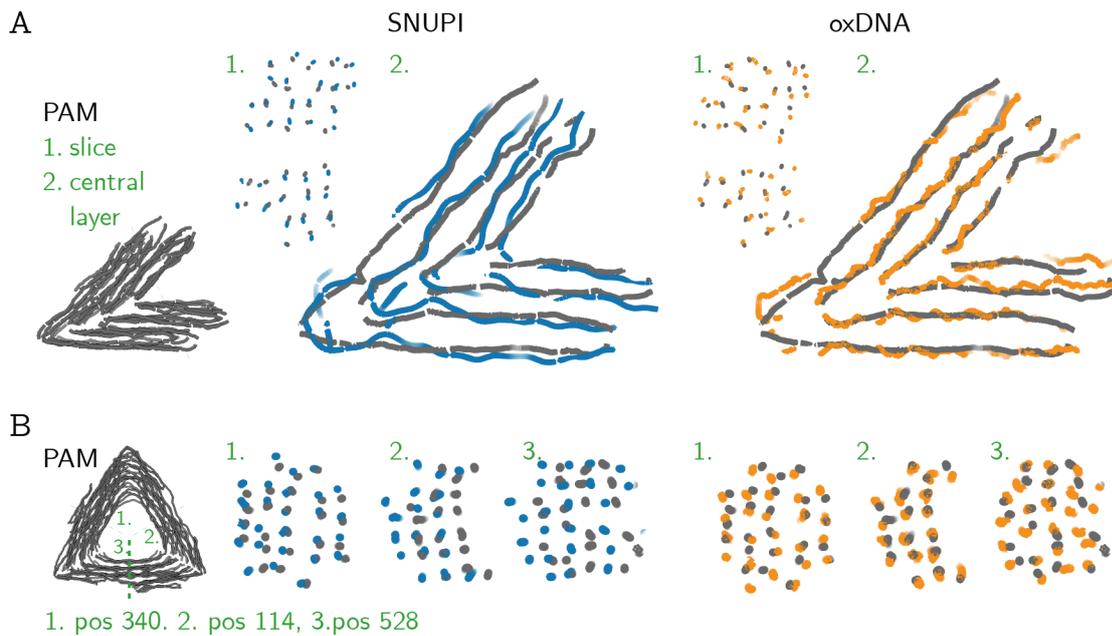
Figure 5.5 | *Coarse prediction for honeycomb-lattice.* *Helical trace comparison of the reference PAM (gray) to SNUPI (blue, middle) and oxDNA (orange, right).* **(A)** *Corner Study Version 3 origami design compared to the PAM by (1.) the cross-section at both ends of the corner arms, (2.) the central helix column of the design diagram.* **(B)** *Helical cross-section for each of the three sides of the B-Triangle object. The three slices are oriented relative to the triangle plane of the PAM (left). Slice markers (green, dashed) and numbering indicate the position of the cross-section relative to the PAM (left).*

Finally, we analyze the capability of the two structure prediction approaches to automatically predict DNA origami designs containing corner motifs [79, 86]. The corner angle for these designs is adjusted by the length mismatch of helical pairs of the corner arms. However, the topology close to the corner and the orientation of the base-pair at its pivot are important factors for the precise engineering of the corner. Modifications of the crossover distribution and helical orientation can cause deviations from the target angle of several degrees.

Using the respective multidomain-alignment strategies, both protocols successfully predicted the target geometry at RMSD values of $16\,\text{Å}$ or better. At $11\,\text{Å}$ and $13\,\text{Å}\,mrDNA/oxDNA$ yields significantly better results compared to the FE method. The cause of the discrepancy for the $45°$ corner of the *Corner Study Version 3* design is a slightly larger opening angle of the *SNUPI* prediction when compared to the PAM (Figure 5.5A). Additionally, the complex topology at the corner, which includes several motifs reliant on stacking interaction, causes prediction artifacts (Appendix Figure B.5B). Nonetheless, the cross-section comparison for the helical trace matches the reference model for both predicted configurations. Angle brace

lengths are within $9\,\text{Å}$ and $15\,\text{Å}$ within the values established for the PAM (Section 4.2) for *oxDNA* and *SNUPI* respectively.

In case of the closed multidomain *B-Triangle* object (Figure 5.5B), corner angles are correct in both cases. For the *SNUPI* prediction, the alignment of the three sides is slightly shifted, causing a larger RMSD value. Slight mismatches for the overall dimensions of each side are the likely reason for the mismatch. The dihedral angles of the triangle sides, which are an essential design goal for the tubule assembly, are predicted with high accuracy in both cases.

5.2.5   Structure Prediction Guidelines

In summary, both *SNUPI* and *oxDNA* are capable of predicting structural features of all ten designs of the experimental validation set. This includes independence of DNA origami lattice, scaffold size, and domain geometry. Both coarse-grained models are capable of simulating structures composed of multiple scaffolds, closed and open topologies, and global twist-deformation.

With simulation times at the timescale of minutes, *SNUPI* has become the most viable prediction tool for iterative DNA origami design. Structure prediction with *SNUPI* can be improved significantly by increasing the electrostatic cutoff to at least $3.75\,\text{Å}$ (or $3\lambda_D$), similar to the value used in the *oxDNA* model. The simulation time penalty factor of up to four is mostly insignificant due to the speed of the simulation in general. Either way, predictions are not fast enough for instant visual feedback, which requires prediction times within milliseconds or at most a few seconds.

Both models slightly underestimate the global twist deformation of honeycomb-lattice objects but can predict the twist of square-lattice designs. Other observed deviations from the global structure are within the resolution limits of the experimental reference data. As the parametrization of both models is dependent on limited high-resolution cryo-EM data, further improvements can certainly be achieved by incorporating the data presented in this work into the model parametrization.

Due to its decisively shorter simulation time and lower hardware demands, *SNUPI* is the best model for rapid iterative structure prediction of DNA origami nanostructures. Average RMSD values around $14\,\text{Å}$ for *SNUPI*, compare favorably to the approximately $11\,\text{Å}$ achieved by *oxDNA*. However, as *SNUPI* is dependent on the DNA origami lattice, non-standard design motifs and lattice-independent properties like base-pair hybridization are not included in the prediction framework. Here, *oxDNA* presents a feasible alternative

for structure prediction. Since the model has been developed independently of DNA origami lattices as a general model for DNA, successful structure prediction does not rely on lattice motifs. Effects like duplex-formation or dynamics of single-stranded DNA are equally realizable as non-lattice DNA nano-objects. In cases where the additional accuracy of *oxDNA* is required, the average runtime of a single day is also not unreasonable. Additionally, *oxDNA* can perform any protocol suitable for an MD simulation. This includes, among others, advanced sampling techniques and the computation of ensemble properties. The *oxDNA* code is developed as open source, enabling custom modification and extension of the code.

5.3   Conclusion

The advent of structure prediction of macromolecules using computer simulation presents an opportunity for the field of DNA nanotechnology. Using prediction-guided iterative design enables the development of more complex and more precisely engineered nanostructures. Nonetheless, the size of DNA origami nanostructures imposes a high computational cost for all-atom simulations of these systems. With the implicit solvation provided by the elastic network of *enrgMD*, this barrier can be overcome. This development facilitates atomic-level structure prediction for megadalton DNA nano-objects. By comparison of *enrgMD* prediction to pseudo-atomic models of six high-resolution cryogenic electron microscopy datasets, we highlight the versatility of the approach. Although solely parametrized on a single square-lattice structure, *enrgMD* generalizes well to both types of layered DNA origami lattices and accurately predicts the overall structure for different staple-breaking rule sets. Apart from its predictive power, the elastic network can be utilized as a relaxation method for explicit solvation simulations, particularly in combination with the multi-resolution framework *mrDNA*. For these reasons, *enrgMD* is an important utility of the shrink-wrap-fitting protocol.

However, the atomistic representation of DNA duplexes with *enrgMD* still necessitates long simulation times. Due to the associated computational cost, this approach is not viable for iterative design. By reducing the complexity of the simulated system via coarse-graining, *oxDNA* and *SNUPI* can provide prediction feedback on the timescale of hours and minutes, respectively. We simulated a test set of ten lattice-based DNA origami structures using the *SNUPI* framework. The base-pair-level finite elements can predict the overall structure with on average 14 Å accuracy in a matter of minutes. Different lattice types, object sizes, and aspect ratios were equally well predicted as multidomain structures. In contrast to *mrDNA* and *oxDNA*, *SNUPI* does not require GPU hardware and is compatible with Windows,

macOS, and Linux. *oxDNA* simulations for the same validation set do produce at least equally good structural agreement with the shrink-wrap-fitted PAM reference, compared to *SNUPI*. The coarse MD model is also more robust toward non-standard design motifs like crossover stacks and single-strand DNA segments. But, prediction with *SNUPI* is about two orders of magnitude faster than *oxDNA*, mitigating the lower predictive power.

For most iterative design workflows, *SNUPI* is the model of choice, due to its rapid prediction and accurate representation of the overall structure and topology of DNA origami designs. In cases incompatible with the *SNUPI* framework or simulations including non-lattice DNA motifs, *oxDNA* presents a viable alternative for prediction-guided iterative design. For computational studies that involve DNA duplex formation or dissociation, or prediction of thermodynamic ensemble averages, *oxDNA* is also the more versatile of the two. Since the code is also available open-source the coarse-grained MD approach presents more flexibility. The model also allows simulation of mixed systems including RNA [123, 124] and protein complexes [122]. The *oxDNA* source code [38, 119, 120] and its associated tools [121, 122] are under active development with novel features added at almost a monthly rate.

Despite the limited atomic-level data available at the time of their development, all three DNA origami models presented here generalize well for a variety of designs. We anticipate, that the high-resolution pseudo-atomic model data presented in this thesis provide a more extensive parametrization set for the improvement of existing or the development of novel prediction models for DNA macromolecules.

Chapter 6

Conclusion & Outlook

In this thesis, we introduced shrink-wrap-fitting, a protocol that allows fast, semi-automated pseudo-atomic model building for lattice-based DNA origami nanostructures. The protocol is a combination of several other methods, coupled to target the characteristics and topology of these objects specifically. At its core, the protocol is based on the MDff protocol [130] that, due to its conceptual simplicity, is applicable for megadalton nano-objects. A cascaded fitting procedure [147] mitigates the effect of local minima caused by the quasi-periodic lattice, while the elastic network of *enrgMD* [31] stabilizes the structure during relaxation and speeds up the dynamics of the system. It makes the fully automatic fitting of structural data into high-resolution cryogenic electron microscopy data of megadalton DNA macromolecules possible.

Supplementing the shrink-wrap protocol, we presented the open-source Python package *dnaFit* which implements its methodology. *dnaFit* enables platform-independent model building for volumetric electron density datasets within 24 hours of computing time on a standard desktop computer. Additionally, we showcased the browser-based *FitViewer* tool, which provides zoning and slicing functionalities. These components are useful for in-depth analysis of these structures and the generation of masks for further data processing. With the extensively documented Python package and Jupyter notebook, anyone can build and analyze quasi-atomic models.

We have constructed several pseudo-atomic models with over 200,000 base-pair coordinates using this technology. Thereby, the total amount of DNA-only coordinates available in the PDB was effectively doubled [32]. Due to its hands-on application, easy availability, and the open-source nature of the package, the method has been quickly adopted by the field.[11, 16, 170] Consequently, we expect an increase in the number of DNA depositions

in PDB in the coming years, further expanding the pool of available structural data on B-DNA.

The functionality of the *dnaFit* package has recently been expanded. It is now capable of fitting multidomain structures, provides a low-resolution mode, and facilitates the fitting of a subset of the structure. With subset fitting, a high-resolution subunit can be modeled and quantified in the context of the full topology. As cryo-EM of DNA nanostructures progresses, especially for the resolution of subunits [32, 99], the need may arise to build models for data with a resolution better than 5 Å. In theory, the MDff protocol is not limited to any resolution range. However, higher resolutions may eventually allow for the removal of the base-pair restraints of the elastic network. Even if more elaborate, and computationally more expensive, fitting schemes [149] should become necessary at higher resolutions, shrink-wrap will remain relevant as a fast and reliable way of creating a starting point for further refinement.

We have further demonstrated how the PAMs can be used to compare and quantify several design aspects with up to nucleotide precision. Slices created using the *FitViewer* notebook properly quantify the twist of square lattice-based structures and highlight how these global deformations impact the planarity of layers along the lattice locally. The zoning and cropping functionality of the tool was used to extract data of specific design motifs like crossover stacks and determine the high-precision placements of functional moieties in an iterative design approach. We identified the effects of co-factor coating and buffer condition changes on both overall dimensions and internal structure of DNA origami nano-objects. The flexibility of the protocol enabled its adaptation and use for electron density data with an overall resolution up to 20 Å. Moreover, we illustrated how this workflow supports the iterative design process by permitting the calibration of design parameters. For instance, we measured bevel-angels of tubule forming triangles, without the positional ambiguity within a volumetric dataset. We used similar workflows to place structure enforcing angle-braces and compared the effects of modification between different design variants. This technique also works well in the presence of unresolved subdomains, strongly varying resolution for different parts of the volumetric dataset, and incorrectly assembled motifs. Additionally, we demonstrated the utilization of PAM coordinates for the assessment of the stereochemical quality of the B-DNA duplexes in terms of base-pair step parameters.

The models generated with the shrink-wrap protocol provide an extensive resource to assess the results of computational structure prediction of DNA origami nanostructures. We used the atomic coordinates fitted to a handful of high-resolution electron density maps to establish a base reference of atomic coordinates. This reference was compared

to structural results simulated using three DNA prediction models. The all-atom model *enrgMD* with the *mrDNA* fast relaxation procedure consistently produced RMSF values below the resolution of the experimental cryo-EM data. Computational cost and the structural restraint due to the elastic network present the main drawbacks of this model. Consequently, domains and motifs that deviate strongly from the idealized lattice model are likely to be predicted with low accuracy. We also compared the results of the two coarse-grained prediction models *oxDNA* and *SNUPI*. Compared to the all-atom approach with trace RMSF values of up to 16 Å and 19 Å, respectively. However, the average trace deviations are within a few ångströms of the cryo-EM resolution. We showed that for iterative design pipelines the reduction in overall predictability is offset by the speedup, especially with average *SNUPI* simulation times of less than 15 minutes. This unmatched speed makes *SNUPI* a valuable rapid prediction tool, especially in the early stages of the iterative design process. Additionally, both tools are capable of mitigating at least one of the disadvantages of *enrgMD*, by providing more flexibility in the design space. Since *oxDNA* is not limited to the simulation of lattice-based DNA origami, the model is capable of predicting deviations from the origami lattice, base-pair dehybridization, and subdomain dynamics.

In addition to establishing the strengths and weaknesses of these prediction models, the availability of multiple pseudo-atomic models provides a broad platform to calibrate and validate future models. We aim to further increase the number of available coordinates and diversify their origin to different categories of structures, lattices, and conditions.

Further developments in high-resolution electron microscopy of DNA origami and pseudo-atomic model-guided data processing [32, 200] will extend the number of available high-resolution DNA origami structures. With this expected increase, the available atomic coordinates of DNA-based structures will grow correspondingly. More experimental data suitable for validation of both atomic-level and coarse-grained models will further enable the improvement of existing, or parametrization of novel prediction approaches. We are convinced that these developments foster a positive feedback loop, with rapid prediction-guided iterative design enabling the engineering of DNA macromolecules of even higher quality and design precision. Structural insights extracted from these nano-objects will likewise improve successive prediction tools. In addition, the increase in high-resolution base coordinates may facilitate the utilization of artificial neural networks for the field of DNA origami. The same goal was recently achieved for proteins for both structure prediction [201, 202] and cryo-EM data identification [203]. Particularly the AlphaFold2 model [202] is considered a milestone achievement in protein structure prediction in particular, but also computational biology in general. Furthermore, high-resolution electron density data will

help facilitate the development of complex design motifs. Given a sufficient library of high-resolution motifs and their respective pseudo-atomic structure, the simplification framework provided by the lattice may no longer be required. Although experimental data with better resolution is required to investigate the sequence dependency of the DNA duplex geometry, a set of high-resolution coordinates presents ample chance for exploratory data analysis.

The methodology and computational tools for computational analysis of lattice-based DNA origami nanostructures presented in this work represent a multidisciplinary effort to improve the level of precision for both the design and validation of these megadalton macromolecules. Development of the software as public and open-source invites adaptation of the codebase to novel challenges and analysis pipelines. The MD-based core of the fitting pipeline makes it suitable for extension and improvement, as better results for the cryo-EM resolutions of DNA nano-objects will be achieved.

Appendix A

Computational Methods

A.1 Cascade Protocol

The relaxation procedure is based on molecular-dynamics flexible fitting (MDff) [130, 145]. The cascade protocol is adopted from cascading MDff [147]. Initial models and elastic networks for the fitting procedure are generated using *enrgMD* [31] and *mrDNA* [116].

A.1.1 Initial Protocol Published in Feigl, Kube, Kohler, et al. (2020):

First, an initial model for a given structure has to be computed and roughly aligned with the volumetric electron density data. The caDNAno2.3 strand diagram [27] and the scaffold DNA sequence are used as input to generate an initial pseudo-atomic model with *enrgMD* [31]. Manual pre-alignment with the electron density data using VMD-1.9.3-MacOSX [166] ensures global optimization of the fitting procedure. Prior to the start of the flexible fitting procedure 4,800 steps of steepest-descent energy-minimization improve restraints and the bonded-neighbor geometry.

All simulations of the fitting cascade are performed in the MD simulation framework NAMD2 [165] version 2.12-Linux using the CHARMM:36 [191, 204] force field for nucleic acids. Intermediate electron density data for the cascade steps and grid-based potentials for flexible fitting are generated using VMD packages mdff-0.5 and volutil-1.3. Simulations were performed in the NVT ensemble with short-ranged Coulomb interactions up to 10 Å, and a switch distance of 8 Å. Smooth truncation of the Lennard-Jones potential was employed and no PME long-range corrections are applied. Simulations were performed in a vacuum with dielectric constant 1 and a thermostat at 300 K using Langevin forces for

all non-hydrogen atoms with damping of 0.1 1/ps.

During the flexible fitting procedure a grid-potential weight ζ of 0.3 kcal/mol is used for the cryo-EM data. For the cascade, a series of eight low-pass filtered volumetric datasets with a sequentially increasing resolution is generated. Low-pass filtering is performed using a Gaussian blur of up to 22 Å. For each step of the cascade 12,000 steps of MDff with a time step of 2 fs are performed. After the first three stages of the cascade, 12,000 steps of regular elastic-network guided MD is performed without the grid-based potential active. This step resolves deformations caused by the proper alignment with the cryo-EM data, which is especially pronounced for poor rigid body docking or multidomain structures.

Now a second cascade is initiated with the final frame of the deformation resolve step as the initial model with 6,000 steps for each map. Throughout the protocol the harmonic bonds constituting the elastic network of *enrgMD*, or a subset thereof, are active. These bonds are differentiated by the relation of their acting atoms, with long-range connecting atoms of different helices, short-range bonds connecting atoms within the same helix, and base-pair bonds connecting the nucleotide pairs connected by hydrogen bonds. For any maps with a global resolution of 14 Å or better, long-range bonds are turned off, as individual helices are discernible. For the final step of the cascade, only base-pair restraining bonds are kept active for 12,000 steps against the original map. Finally, the weight on the EM density is increased to 1.0 kcal/mol for 18,000 energy-minimization steps to equilibrate bonded-neighbor restraints of the force field.

After performing the flexible fitting protocol, final coordinate files (.pdb and .mmCIF) are generated. This post-processing includes the computation of a masked electron-density map with a masking distance of 4 Å that can be used for computing the masked cross-correlation.

A.1.2   Protocol improvements for *dnaFit-1.0*:

Instead of generation with *enrgMD* and subsequent energy-minimization, the initial model is generated using the multi-resolution structure prediction framework *mrDNA* [116]. Since the initial model will be significantly closer to the solution conformation, cascade 6,000 steps per cascade map are sufficient for most structures. In addition to relaxing the structure without the presence of a grid-based potential, an annealing step is added in between the first and second cascade. Here the system is gradually heated by 100 K for 12,000 steps, before stepwise cooling to the original temperature. Throughout this procedure, the grid-based potential is active with a reduced weight of 0.01 kcal/mol for MDff. The complete

post-processing is performed automatically by *dnaFit* with an additional generation of a human-readable version of the design- pseudo-atomic model linkage.

By default, all single-stranded residues are excluded from the grid-based potential by generating a custom *grid.pdb* file in the context of the caDNAno design file. The subdomain fitting pipeline is accessible by supplying a *grid.pdb* file instead of default automatic generation.

Low-Resolution Fitting Protocol: For electron density data with a resolution of 15 Å or worse, application of the complete protocol is not warranted as base-pair level detail and helical grooving are not discernible. However, the placement of elastic network retrained atomic models at the position of helices can be used to approximately annotate the volumetric dataset. DNA origami design motifs like crossovers are distinguishable at these resolutions allowing validation of proper global alignment of the helices. For this purpose the fitting protocol can be followed up to the point of removal of intra-helical elastic network constraints, directly followed by the minimization step and post-processing.

A.2 Software Availability

All code described here is available on GitHub and is part of the public domain. Release numbers listed below reflect the version at the time of publication of this work.

The code described in the following section was developed and tested on Unix-like systems, specifically CentOS Linux and Mac OS X. The Python packages are developed with Python 3.8, the Jupyter Notebook was tested in Safari but is expected to work in other environments. Full disclosure of license information and contributions to the code is available on GitHub. In general, we advise the use of isolated virtual environments for the installation of the Python packages.

If you encounter any issues, please check out the latest release and use the respective GitHub-Issue page for reporting errors.

A.2.1 Shrink-Wrap Fitting Package

Package-identifier: **dnaFit**
URL: `https://github.com/elija-feigl/dnaFit`
License: *GNU Public License Version 3*
Version: 1.0.1.alpha

Description: A Python 3 package that contains a command-line of scripts related to the fitting of pseudo-atomic models into cryo-EM maps of DNA-Origami structures. Implements the *Shrink-Wrap Fitting* (*Multi-Resolution Elastic Network-Guided Cascaded Molecular Dynamics Flexible Fitting*) pseudo-atomic model building protocol.

Requirements:

- **Python $\geq$ 3.8**
    - numpy $\geq$ 1.21.0
    - click $\geq$ 8.0.1
    - MDAnalysis $\geq$ 1.1.1
    - mrcfile $\geq$ 1.3.0
    - nanodesign $\geq$ 3.0.0. (@ https://github.com/elija-feigl/nanodesign)
    - pdb2cif $\geq$ 1.0.0 (@ https://github.com/elija-feigl/pdb2cif)

Installation: To install *dnaFit* either download and unpack the code from GitHub before installing the package with a package managing frontend like *pip*:

```
pip install .
```

Or use the frontend to install directly from GitHub:

```
pip install git+https://github.com/elija-feigl/dnaFit#egg
    =dnaFit
```

Usage: After installation, the package also serves as the entry point for the command-line interface. To display the documentation of the application execute it by name in the shell.

```
dnaFit --help
```

```
Usage: dnaFit [OPTIONS] COMMAND [ARGS]...
    Cascaded \textit{mrDNA}-driven MD flexible fitting script
        module:
    dnaFit commands:
        main pipeline:
        1. mrDNA:  \textit{mrDNA} structure prediction with
            custom settings for caDNAno file. (includes the
            creation of prep folder for fitting)
        2. vmd_info:   print info for rigid body docking
            with VMD
```

```
    3. fit:      shrink wrap fitting if of rigid body
       docked \textit{mrDNA} prediction. (includes mask,
       pdb2cif, and link)


Options:
    -v, --version  Show __version__ and exit.
    -h, --help     Show this message and exit.


Commands:
    center-on-map  Recenter atomic model center-of-mass on
       MRC cryo-EM map center and write a structure (.pdb)
       file.
    fit            Cascaded \textit{mrDNA}-driven MD flexible
        fitting (shrink wrap fitting) to MRC cryo data.
       Creates \textit{dnaFit} folder. This folder will
       contain the final results \textit{prefix-last.cif}.
    link           Links structural information of the
       CADNANO design file to fitted atomic model files TOP,
       CONF.
    mask           Creates a masked MRC map of voxels
       occupied by the PAM.
    mrDNA          \textit{mrDNA} simulation of CADNANO
       design file with custom settings followed by
       preparation of files for "fit" command.
    pdb2cif        Generate atomic model in mmCIF format from
        NAMD structure files using the pdb2cif package
    prep           Prepare \textit{mrDNA} results for "fit"
       command.
    vmd-info       Print VMD command for rotation of
       structure file (.pdb) around its center of mass.
```

Reference: When using the *dnaFit* Python Package in published work, please cite the following paper:

Feigl E., Kube M., Kohler F., et al. Revealing the structures of megadalton-scale DNA complexes with nucleotide resolution. Nat Commun 11, 6229 (2020). `https://doi.org/10.1038/s41467-020-20020-7` [32]

A.2.2   FitViewer Notebook

Identifier: **FitViewer.ipynb**
URL: `https://github.com/elija-feigl/FitViewer`
License: *GNU Public License Version 3*
Version: 1.3.0

Description: Interactive Jupyter Notebook for context-based zoning and cropping of cryo-EM maps using pseudo-atomic models fitting into cryo-EM electron density maps of lattice-based DNA origami nanostructures.

Requirements: FitViewer requires an iPython kernel with the following dependencies:

- **Python $\geq$ 3.8**
  - dnaFit $\geq$ 1.0.0 (@ https://github.com/elija-feigl/dnaFit)
  - jupyter notebook $\geq$ 6.4.8
  - iPython $\geq$ 8.0.0
  - ipywidgets $\geq$ 7.6.5

Usage: A Jupyter Notebook can be started from an appropriate editor or in the browser via shell:

```
Jupyter Notebook FitViewer.ipynb
```

Instructions on the use of the notebook are directly incorporated via mark-up cells.

Reference: When using *dnaFit* Python Package in published work, please cite the following paper:

Feigl E., Kube M., Kohler F., et al. Revealing the structures of megadalton-scale DNA complexes with nucleotide resolution. Nat Commun 11, 6229 (2020). `https://doi.org/10.1038/s41467-020-20020-7` [32]

A.2.3   PDB to mmCif Format Conversion

Package-Identifier: **pdb2cif**
URL: `https://github.com/elija-feigl/pdb2cif`
License: *GNU Public License Version 3*

Version: 1.0.4

Description: This tool converts protein databank files (.pdb) into mmCIF (.cif) format for upload to the PDB [158] and use in chimera [175] and ChimeraX [205].

Current supported (.pdb) sources:

- *mrDNA*
- *tacoxDNA*
- *SNUPI*

Requirements:

- **Python $\geq$ 3.8**
  - numpy $\geq$ 1.21.0
  - click $\geq$ 8.0.1

Installation: To install *pdb2cif* either download and unpack the code from GitHub before installing the package with a package managing frontend like *pip*:

```
pip install .
```

Or use the frontend to install directly from GitHub:

```
pip install git+https://github.com/elija-feigl/pdb2cif#
   egg=pdb2cif
```

Usage: Start the command-line interface by typing the name of the package into the shell.

```
pdb2cif --help
```

```
Usage: pdb2cif [OPTIONS] PDB
  Generate atomic model in mmCIF format from NAMD structure
     file (.pdb).
  PDB is the name of the NAMD configuration file [.pdb]

Options:
  -h, --help     Show this message and exit.
  -v, --version  Show __version__ and exit.
```

```
--remove-H      Remove hydrogen atoms.
--alt-chain     Use alternative chain identifier.
--snupi         Convert from \textit{SNUPI} pdb.
--flip-fields   Flip the values of occupancy and temperature
      fields.
```

A.2.4   Nanodesign

Package-identifier: **nanodesign**
URL: `https://github.com/elija-feigl/nanodesign`
License: *Apache-2.0 license*
Version: 1.0

Description: A Python 3 port of the currently discontinued *nanodesign* package `https://github.com/Autodesk/nanodesign`. The package provides a development framework for DNA nanostructures. *Nanodesign* offers read and write routines for common DNA origami design software, including *cadnano* and creates a dynamic data structure to store information about the designed macromolecule.

A.3   DnaFit Code Structure

In this section, we describe the overall structure of the *dnaFit* Python package and detail the functionality of the most important routines. The package itself acts as a backend for the FitViewer application and automatized analysis protocols, but also presents a command line interface for administering the shrink-wrap fitting protocol. The command line interface is constructed using the Python module click. The package is split into seven submodules, listed in alphabetical order.

- *core*: This module contains the *mrDNA* initial-model generation wrapper, with two sets of custom parameters for single- and multidomain DNA origami designs. General utility functions, model parameters, and constants are also stored here.
- *data*: Class definition for real-life equivalent objects are located in this module.
- *fit*: Here, the shrink-wrap protocol is implemented and encapsulated in the Cascade Python class.
- *link*: Linkage between the atomic structure file and the caDNAno strand diagram is processed within this module.

- *resources*: NAMD2 force field files, as well as input file templates for NAMD2 and VMD are located in this folder.
- *scripts*: This module serves as the entry point for commands defined for the CLI.
- *viewer*: Functionality exclusively utilized for the FitViewer application is collected in this module. It Implements various subset selection methods (via widget or method call) and output routines for modified volumetric and atomic structure files.

In general, the code base is highly encapsulated following an object-oriented programming paradigm. Consequently, any utility is bound to an abstract object, the most important of which are discussed in the following.

Data Module

The main objective of the *dnaFit* packages is the creation of a contextual connection between the topology of the caDNAno strand diagram and the geometry of the electron-density data fitted by the atomic model. Consequently, the two main data classes implement these concepts: the design and the pseudo-atomic model.

The *Design* class contains the complete information on the DNA origami design. Parsing of the caDNAno strand diagram is performed with the nanodesign Python package A.2.4. The existing data structure of nanodesign is subsequently extended by containers for scaffold and staple strands, as well as a lookup table for identifier and position of all bases of the system. Additionally, the strand and helix parsing order is resolved to match the pseudo-atomic model. The *Fit* class parses and stores information on the PAM via the MDAnalysis package.

In addition, both these concepts represent a description of DNA, either in terms of its design or its experimental readout. Since this context is essential for quantitative analysis of DNA, we implement classes for B-DNA duplexes, hydrogen-bonded base-pairs, and individual bases. The *Base* and *Basepair* class define the coordinate reference frames and identities of each nucleotide. These coordinates are then used to compute the properties of a B-DNA duplex (*BDna*). The current release implement routines for the computation of base-pair step properties twist, rise, tilt, roll, shift, and slide. Trace data is computed based on the origin of the base-pair reference frame. Evaluation of base step properties, crossover angles, and backbone dihedral angels is currently in development.

Fit Module

The main protocol of shrink-wrap fitting is implemented in this module. Implementation of the model building workflow is encapsulated in the *Cascade* class with the parameter set stored separately in the *CascadeData* class. First, with subprocess calls to VMD, the low-pass filtered cascade maps and MDff files are generated. Then, each fitting stage is executed via NAMD2, before the final processing of the simulation data is performed with VMD. Finally, the results of the MD simulation are parsed and processed with *dnaFit* to mmCIF structure files for PDB upload and masked electron density data for cross-correlation validation. Additionally, design-to-PAM linkage information is computed and stored in a human-readable format.

Link Module

The *linker* class matches DNA strands of PAM and caDNAno strand diagram based on the topology parsing rules of *mrDNA* and *enrgMD*. Once the correct counterpart of each strand is identified, the base identifier can be matched from the DNA origami design to the structural model. The identifier pairs are subsequently utilized to enrich the PAM with the topological and lattice-specific information of the design. Among others, base-pairs, crossover and nick location, staple color, and lattice position (helix number & base position) are coordinated with the residue identifiers of the model. The resulting lookup tables are stored as dictionaries inside a specialized *linkage* class. Both analysis pipelines based on *dnaFit* and the *FitViewer* Jupyter notebook use the *linkage* class for contextualizing the experimental data. The class also implements the human-readable output method for the linkage information.

A.4   Structure Prediction Protocols

A.4.1   *enrgMD* Parameters:

Initial models for *enrgMD* structure prediction were generated using the web server [31] and the *enrgmd*-tool provided by the *mrDNA* [116] package. All simulations are performed in NAMD2 [165] version 2.12-Linux using the CHARMM:36 [191, 204] force field for nucleic acids. The simulations were performed in the NVT ensemble with short-ranged Coulomb interactions up to $10\,\text{Å}$, and a switch distance of $8\,\text{Å}$. Smooth truncation of the Lennard-Jones potential was employed and no PME long-range corrections are applied. Simulations were performed in a vacuum with dielectric constant 1 and a thermostat at $300\,\text{K}$ using Langevin forces for all non-hydrogen atoms with damping of $0.1\,1/\text{ps}$.

A.4.2 *oxDNA* Protocol:

All *oxDNA* simulations presented in this work utilize *mrDNA* [116] (v1.0a.dev53) for the relaxation phase. The coarse-grained simulations use model version *oxDNA*2 [76], with *oxDNA*-code version (v2.4). This drastically decreases simulation times and enables the fully autonomous simulation of multidomain structures. Simulations in *mrDNA* are performed at 295 K with 10e7 coarse steps and 10e7 fine steps followed by a regular *oxDNA* simulation. Sequence information was supplied via the sequence of the scaffold strand with undefined single-stranded staple bases set as thymidine. For the multidomain *B-Triangle* structures, a coarse-bond-cutoff of 600 nm was utilized to enable subdomain alignment.

The subsequent *oxDNA* relaxation consists of 10e3 steps of energy minimization to reduce artifacts in the initial configuration and 10e6 regular MD steps for the final equilibration, followed by 9e6 steps production with a time step of 0.003 fs. The temperature is kept at 295 K using the *john* thermostat with a diffusion coefficient of 2.5. The overall salt concentration was set to 20 M monovalent ions.

The average ensemble configuration used as the structure prediction result was generated using the *oxDNA* Analysis Toolbox [122] and converted to PDB format using *tacoxDNA* [121].

A.4.3 *SNUPI* Protocol:

Finite Element prediction using *SNUPI*_2.01-Mac, was executed with the MacOS command line version. Simulations were executed with the default setting of 30 time steps at 300 K and 20 mM $MgCl_2$. Staple sequences were supplied via caDNAno. The cutoff distance for electrostatic interactions was increased from the default of 2.5 nm to 5.0 nm, corresponding to approximately four times [39] the Debye length at 20 mM $MgCl_2$. While the standard settings provide accurate information about the overall shape of the structure, the increased cutoff value improved the prediction of absolute overall dimensions, resulting in a better RMSD (see B.3) Gaussian sampling of single-stranded DNA properties was enabled for all simulations

For multidomain structures, the partition & relocation framework implemented in *SNUPI*_2 [128] was utilized. Results of each structure were saved in PDB format, which was used for helical trace analysis.

Appendix B

Supporting Information

B.1   Pseudo-Atomic Model Building

For the validation of the model building protocol, PAMs of two independent subsets of particles from the complete cryo-EM dataset are generated (Figure B.1A). As independent subsets, the two half-maps 1 & 2 produced during RELION [97, 98] post-processing are used. For each of the two half-map PAMs as well as the full model, a simulated electron density map is created at the reported resolution (PAM-map). Using the Fourier Shell Correlation (FSC) the three simulated PAM-maps can be compared to the experimental data. First, the FSC of the full cryo-EM map and the final fitted model PAM-map (FSC_full) are computed. Secondly, PAM-map simulated for a half-map 1 is compared to its respective cryo-EM half-map 1(FSC_work). Finally, PAM-maps of half-map 1 are measured against the other, independent half-map 2 (FSC_test). If strong deviations of FSC_work and FSC_test are observable, this indicates potential overfitting of the experimental dataset by the model building protocol (Figure B.1B). The FSC curves generated with half-map data should also feature the same characteristics as FSC_full, yet at a lower resolution. Since the two half-map FSCs do not differ significantly, shrink-wrap fitting is viable for this dataset ad can be performed on the high-resolution cryo-EM data (Figure B.1C).

	PDB-id	Resolution [Å]	mCCC [a.u.]
Pointer Version 1	—	12.0	0.916
Pointer Version 2	7ARE	7.4	0.874
Twisttower Native Twist	7ARV	7.0	0.927
Twisttower Corrected Twist	7ARY	8.1	0.941
Whopper	—	7.3	0.867
Whopper Version 2	—	7.6	0.874
Whopper Junior	—	8.6	0.896
16-Helix Bundle	7ARQ	10.0	0.940
48-Helix Brick	7ART	9.9	0.932
126-Helix Bundle	7AS5	8.8	0.909
B-Brick	7NPN	11.6	0.951
A-Brick	—	10.0	0.878
Triplex Bullet	—	7.4	0.895
Triplex Bullet (triplexes formed)	—	7.5	0.883
A-Triangle	—	20.0	0.938
B-Triangle	—	24.1	0.948
X-Triangle	—	23.5	0.923
V-Triangle	—	23.5	0.909
Corner Study Version 3	—	8.2	0.933
Corner Study Version 6	—	7.1	0.931

Table B.1 | ***Pseudo-atomic model references.*** *Protein Data Bank identifiers for published high-resolution models reported overall cryo-EM resolution and masked cross-correlation coefficients. Experimental resolution estimates are used as input for the model building protocol. Values for the mCCC were computed using the VMD package mdff-0.5.*

	Resolution [Å]	mCCC [a.u.]
B-Brick (bare Fob5)	11.6	0.951
B-Brick K10 0.5 Fob5	16.4	0.952
B-Brick K10-PEG 0.75 Fob5	12.6	0.925
B-Brick K10-PEG 0.75 PBS	15.1	0.911
B-Brick K10-PEG 0.2 Fob5	12.0	0.933
B-Brick K10-PEG 100 Fob5	16.6	0.896
B-Brick K10-PEG 100 PBS	18.8	0.884

Table B.2 | ***B-Brick pseudo-atomic model references:*** *Reported overall cryo-EM resolution and masked cross-correlation coefficients of the B-Brick for various coatings and buffer conditions. Experimental resolution estimates are used as input for the model building protocol. Values for the mCCC were computed using the VMD package mdff-0.5.*
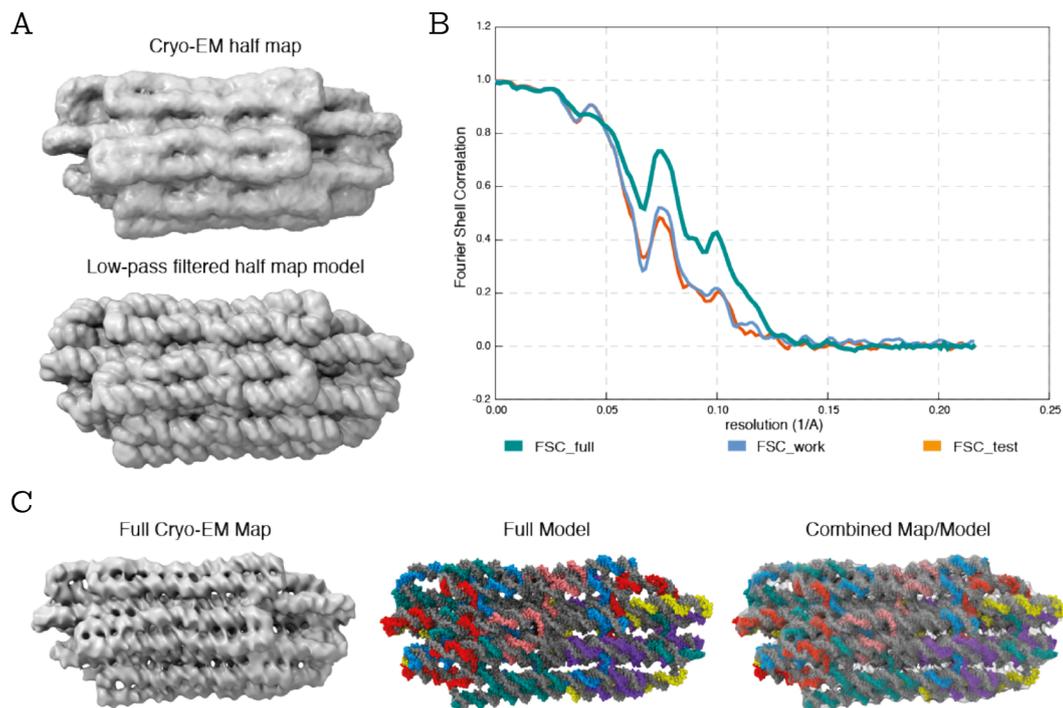
Figure B.1 | _Fourier Shell Correlation (FSC) validation._ (A) _Cryo-EM half-map and simulated map for a PAM build for this data subset._ **(B)** _Comparison of the FSC of the full cryo-EM map and the final PAM (FSC_full), half-map 1 to PAM-map 1 (FSC_work), and half-map 1 to PAM-map 2 (FSC_test)._ **(C)** _Final results of the model building protocol for the complete experimental data (left) depicted as model (middle) and map-model overlay (right)._
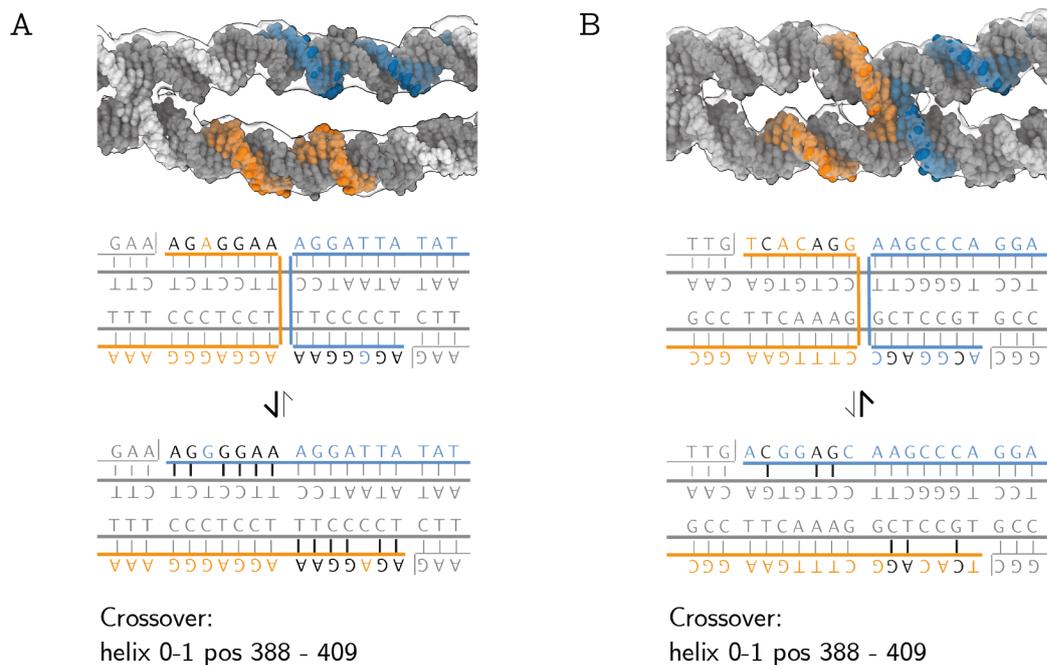
Figure B.2 | _Triplex Bullet._ (A) _Triplex Bullet (triplexes formed): Sequence homology of a single asymmetric crossover (top). The intended crossover (middle) does not appear in the ensemble, with the alternative conformation (bottom) occurring exclusively in the experimental data (also 4.4D). Segments display a matching sequence of 6 out of 7 bases._ **(B)** _Triplex Bullet (sequence variation): As in (A), but without a sequence homology at the asymmetric crossover, resulting in a properly formed crossover._

B.2 B-DNA Duplex Property Computation

Step	Twist [°]	Roll [°]	Tilt [°]	Rise [Å]	Shift [Å]	Slide [Å]
AA	36.73 ±3.49	37.03 ±3.73	12.40 ±7.81	3.28 ±0.53	-0.90 ±0.77	0.12 ±0.42
AC	35.84 ±4.27	36.61 ±4.34	11.14 ±7.09	3.30 ±0.49	-0.98 ±0.66	0.20 ±0.40
AG	36.32 ±4.23	36.38 ±4.65	10.55 ±6.81	3.40 ±0.46	-0.82 ±0.69	0.18 ±0.45
AT	35.87 ±4.43	36.35 ±4.71	11.31 ±8.00	3.43 ±0.51	-0.71 ±0.70	0.12 ±0.45
CA	36.12 ±4.13	36.64 ±4.18	9.59 ±6.99	3.46 ±0.55	-0.58 ±0.65	0.14 ±0.43
GA	36.10 ±4.11	36.72 ±4.15	10.62 ±7.43	3.36 ±0.52	-0.73 ±0.71	0.20 ±0.41
GC	35.78 ±4.75	36.83 ±4.48	10.14 ±8.88	3.32 ±0.50	-0.78 ±0.68	0.11 ±0.40
GG	37.82 ±4.09	37.99 ±4.78	9.50 ±6.35	3.38 ±0.49	-0.64 ±0.75	0.28 ±0.41
–	36.37 ±4.12	36.81 ±4.34	10.82 ±7.46	3.36 ±0.51	-0.78 ±0.72	0.17 ±0.43

Table B.3 | *Base-pair step properties.* *Sequence-dependent base-pair step parameters (three rotational, three translational) and the overall average of enrgMD prediction. Average values are computed using a uniform base-step distribution. Reported error values represent the spread calculated assuming Gaussian distributions. Data extracted from Pointer Version 2, both Twisttower variants, and the 126-Helix Bundle designs.*

B.2.1 Helical Trace Calculation

The linker class of the Python package *dnaFit* originally matches DNA strands of PAM and caDNAno strand diagram based on the topology parsing rules of *mrDNA* and *enrgMD*. As both tac*oxDNA* and *SNUPI* use different parsing rules, an alternate strand matching module was developed. This variation of the linkage-routine matches strand identifiers solely based on their sequence. Using pattern matching principles, this approach is also robust against random sequence assignment of single-stranded domains and deletion of terminal single-stranded segments. Since atomic structure file parsing is performed using the MDAnalysis package [176], no NAMD2 [165] structure file (.psf) is required as input. *dnaFit* can automatically determine the appropriate strand matching algorithm. Importantly, only a subset of functionalities of *dnaFit* is available for structure files generated with *SNUPI* and *oxDNA* structure prediction tools.

Base-pair coordinate-frames and midpoint are computed based on the standard established by Dickerson et al. [64, 65] and depicted in Figure B.3. For each pair, a right-handed coordinate frame is placed at its midpoint such that the x-axis points towards the major groove and the y-axis along the C1'-C1' vector in sequence-strand direction. The z-axis is defined by the right-handed rule resulting in 5'-3' directionality for B-DNA. The direction of the x-axis is defined by the pseudo-dyad of the C1'-C1' vector within the base plane.

The helical origin is positioned at the intersection of this pseudo-dyad and the vector from the C6 atom of the pyrimidine base to the C8 atom of the purine nucleobase.
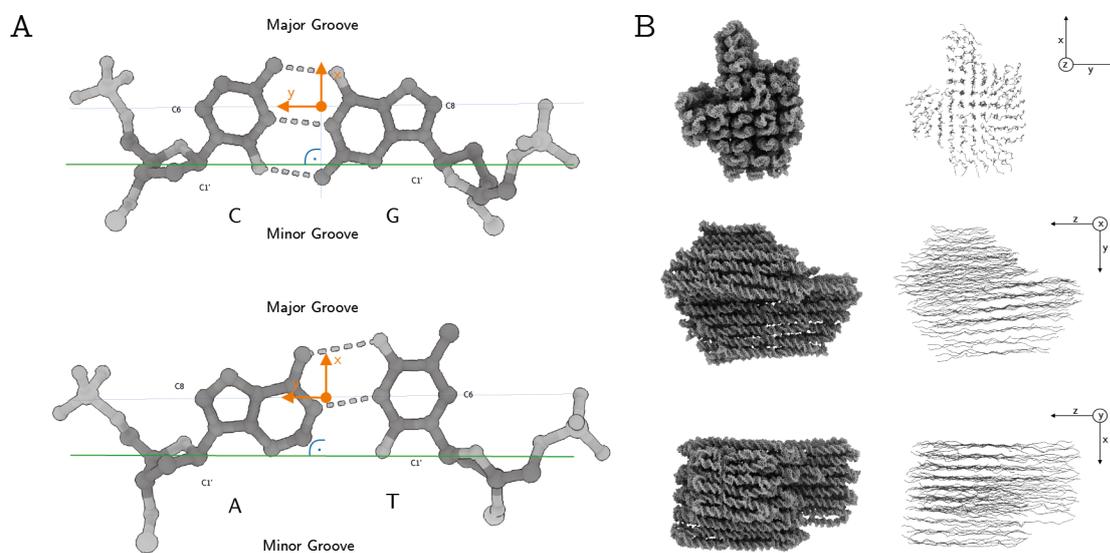

**Figure B.3 | *Base-pair reference frame and helical trace. (A)* *Idealized base-pair reference frame of a purine (R) to pyrimidine pair. (B) Pseudo-atomic model (left) and helical trace (right) computed from the base-pair midpoints the Pointer Version 2.*

B.3   Structure Prediction

[min]	oxDNA	SNUPI [default]	SNUPI
Pointer Version 2	26 h 36	4.5	13.8
Twisttower Native Twist	45 h 19	11.9	26.8
Twisttower Corrected Twist	31 h 43	5.5	11.9
Whopper	19 h 34	0.7	2.4
Whopper Junior	9 h 47	0.2	0.5
16-Helix Bundle	15 h 18	0.2	0.6
48-Helix Brick	39 h 6	5.2	12.2
126-Helix Bundle	84 h 12	76	157
Corner Study Version 3	17 h 42	2.5	14.0
B-Triangle	33 h 43	3.8	15.1

Table B.4 | Simulation Time. *Single-execution timing data for coarse structure prediction models. SNUPI simulations were performed on a single personal computer (Intel i5–8279U 2.4 GHz 4-core CPU). Simulations with mrDNA and oxDNA were performed on a single GPU (NVIDIA GeForce GTX 1080)*

[Å]	oxDNA	$SNUPI^{default}$	$SNUPI^{2R_{elec.}}$
Pointer Version 2	6.89	13.31	7.91
Twisttower Native Twist	14.02	22.73	11.02
Twisttower Corrected Twist	11.99	14.78	12.48
Whopper	10.03	12.81	10.06
Whopper Junior	8.04	11.54	9.35
16-Helix Bundle	9.37	11.48	11.14
48-Helix Brick	15.43	16.95	18.94
126-Helix Bundle	11.25	19.90	11.25
Corner Study Version 3	10.71	14.80	15.51
B-Triangle	13.23	19.12	15.92

Table B.5 | Trace RMSD: Coarse-Grained Models. *Helical trace RMSD for the mrDNA/oxDNA (version 2) prediction protocol, SNUPI version 2.1 with default settings, and with doubled electrostatic cutoff for a subset of ten DNA origami models. Trace data of the pseudo-atomic models of each model were used as reference models.*
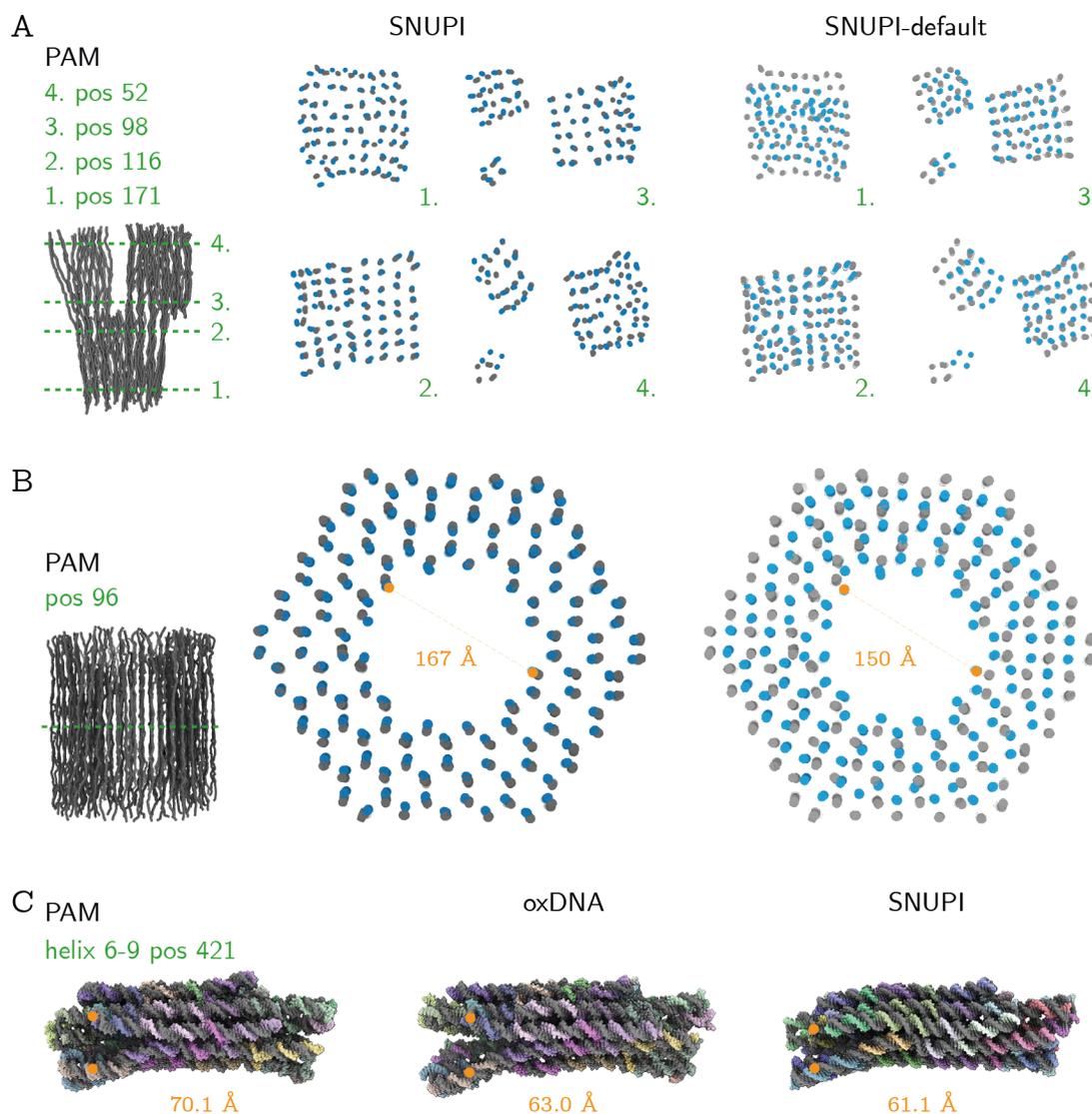
Figure B.4 | *SNUPI electrostatic cutoff comparison.* *Comparison of the cross-sectional slices of the reference PAM (gray, left) to SNUPI with modified (middle, blue) and default settings (right, light blue).* ***(A)*** *Twisttower Native Twist predictions are displayed at four positions that mark both ends of the four subdomains. Marked slices mark both ends of the 8×8 domain (caDNAno position 171 & 116) and small domains (caDNAno position 98 & 52)* ***(B)*** *126-Helix Bundle object sliced in the center around position 96 of the caDNAno diagram.* ***(C)*** *16-Helix Bundle design with flaring ends of helix pair 6-9. Helical trace distance estimates for both prediction methods are within $10\,\text{Å}$ of the reference PAM (left). Slice markers (green, dashed) indicate the position of the cross-section relative to the PAM (left).*
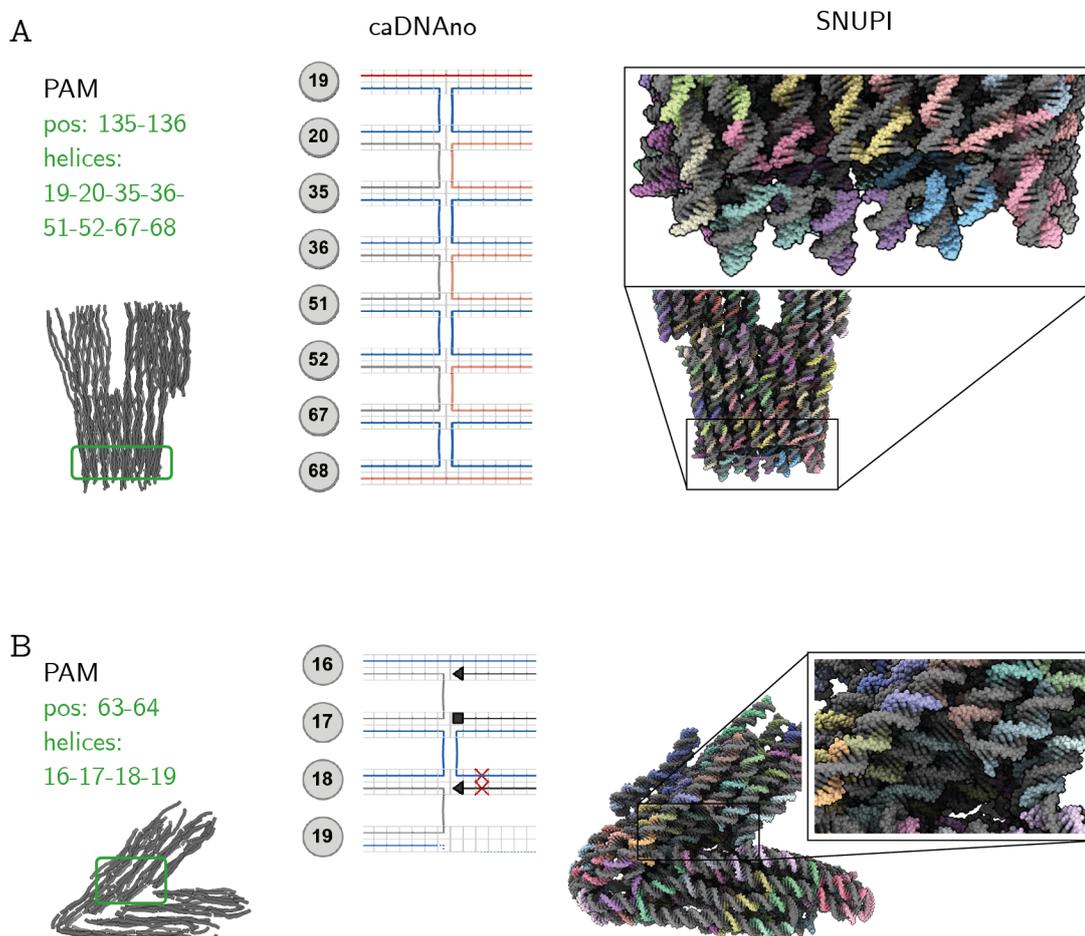
Figure B.5 | *SNUPI crossover-stack artifacts.* *Illustration of the incorrectly depicted crossover-stack motifs with the SNUPI structure prediction framework. Stacking interactions within a DNA origami design are not modeled as finite elements. Markers (green, dashed) indicate the location relative to the PAM (gray, left).* **(A)** *Eight-helix crossover-stack (middle) of the Twitter NativeTwist spanning the entire side of the 8×8 domain. The two stacking interfaces cause an opening in the structure resulting from the absence of covalent bonds in the strand diagram (right).* **(B)** *Six-helix crossover-stack (middle) of the Corner Study Version 3 object. As the lower end of the stack is open, due to a nick in the staple strand diagram (right), the helices can detach from their lattice position, creating a simulation artifact.*

Bibliography

[1] Paul W. K. Rothemund. "Folding DNA to create nanoscale shapes and patterns". In: *Nature* 440.7082 (Mar. 2006), pp. 297–302.

[2] Klaus F. Wagenbauer, Christian Sigl, and Hendrik Dietz. "Gigadalton-scale shape-programmable DNA assemblies". In: *Nature* 552.7683 (2017), pp. 78–83.

[3] Shawn M. Douglas, Ido Bachelet, and George M. Church. "A logic-gated nanorobot for targeted transport of molecular payloads". In: *Science (80-.).* 335.6070 (2012), pp. 831–834.

[4] Thomas G. Martin et al. "Design of a molecular support for cryo-EM structure determination". In: *Proc. Natl. Acad. Sci. U. S. A.* 113.47 (Nov. 2016), E7456–E7463.

[5] Jonas J. Funke and Hendrik Dietz. "Placing molecules with Bohr radius resolution using DNA origami". In: *Nat. Nanotechnol.* 11.1 (Jan. 2016), pp. 47–52.

[6] Fei Zhang et al. "Complex wireframe DNA origami nanostructures with multi-arm junction vertices". In: *Nat. Nanotechnol.* 10.9 (2015), pp. 779–784.

[7] Erik Benson et al. "DNA rendering of polyhedral meshes at the nanoscale". In: *Nature* 523.7561 (2015), pp. 441–444.

[8] Erik Benson et al. "Evolutionary Refinement of DNA Nanostructures Using Coarse-Grained Molecular Dynamics Simulations". In: *ACS Nano* 13.11 (2019), pp. 12591–12598.

[9] Eva Bertosin et al. "A nanoscale reciprocating rotary mechanism with coordinated mobility control". In: *Nat. Commun.* 12.1 (Dec. 2021), p. 7138.

[10] Thomas Gerling et al. "Dynamic DNA devices and assemblies formed by shape-complementary, non-base pairing 3D components". In: *Science (80-.).* 347.6229 (Mar. 2015), pp. 1446–1452.

[11] Anna-katharina Pumm et al. "A DNA origami rotary ratchet motor". In: *Nature* 607.7919 (July 2022), pp. 492–498.

[12]   Hendrik Dietz, Shawn M. Douglas, and William M. Shih. "Folding DNA into twisted and curved nanoscale shapes". In: *Science (80-.).* 325.5941 (Aug. 2009), pp. 725–730.

[13]   Christian Sigl et al. "Programmable icosahedral shell system for virus trapping". In: *Nat. Mater.* 20.9 (2021), pp. 1281–1289.

[14]   Tural Aksel et al. "Molecular goniometers for single-particle cryo-electron microscopy of DNA-binding proteins". In: *Nat. Biotechnol.* 39.3 (Mar. 2021), pp. 378–386.

[15]   Nandhini Ponnuswamy et al. "Oligolysine-based coating protects DNA nanostructures from low-salt denaturation and nuclease degradation". In: *Nat. Commun.* 8.May (2017), pp. 1–9.

[16]   Eva Bertosin et al. "Cryo-Electron Microscopy and Mass Analysis of Oligolysine-Coated DNA Nanostructures". In: *ACS Nano* 15.6 (June 2021), pp. 9391–9403.

[17]   Thomas Gerling et al. "Sequence-programmable covalent bonding of designed DNA assemblies". In: *Sci. Adv.* 4.8 (2018), pp. 1–12.

[18]   Pierre Stömmer et al. "A synthetic tubular molecular transport system". In: *Nat. Commun.* 12.1 (Dec. 2021), p. 4393.

[19]   Philip Ketterer, Elena M. Willner, and Hendrik Dietz. "Life Sciences: Nanoscale rotary apparatus formed from tight-fitting 3D DNA components". In: *Sci. Adv.* 2.2 (Feb. 2016), pp. 1–8.

[20]   Hongzhou Gu et al. "A proximity-based programmable DNA nanoscale assembly line". In: *Nature* 465.7295 (May 2010), pp. 202–205.

[21]   Rémi Veneziano et al. "Role of nanoscale antigen organization on B-cell activation probed using DNA origami". In: *Nat. Nanotechnol.* 15.8 (2020), pp. 716–723.

[22]   Qian Zhang et al. "DNA origami as an in vivo drug delivery vehicle for cancer therapy". In: *ACS Nano* 8.7 (2014), pp. 6633–6643.

[23]   Anton Kuzyk et al. "DNA-based self-assembly of chiral plasmonic nanostructures with tailored optical response". In: *Nature* 483.7389 (Mar. 2012), pp. 311–314.

[24]   Chao Zhou, Xiaoyang Duan, and Na Liu. "A plasmonic nanorod that walks on DNA origami". In: *Nat. Commun.* 6.1 (Nov. 2015), p. 8102.

[25]   Mihir Dass et al. "DNA Origami-Enabled Plasmonic Sensing". In: *J. Phys. Chem. C* 125.11 (Mar. 2021), pp. 5969–5981.

[26]   Carlos Ernesto Castro et al. "A primer to scaffolded DNA origami". In: *Nat. Methods* 8.3 (Mar. 2011), pp. 221–229.

[27]   Shawn M. Douglas et al. "Self-assembly of DNA into nanoscale three-dimensional shapes". In: *Nature* 459.7245 (May 2009), pp. 414–418.

[28] J. D. WATSON and F. H. C. CRICK. "Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid". In: *Nature* 171.4356 (Apr. 1953), pp. 737–738.

[29] Richard Wing et al. "Crystal structure analysis of a complete turn of B-DNA". In: *Nature* 287.5784 (1980), pp. 755–758.

[30] Paul J. Hagerman. "Flexibility of DNA". In: *Annu. Rev. Biophys. Biophys. Chem.* 17.1 (June 1988), pp. 265–286.

[31] Christopher Maffeo, Jejoong Yoo, and Aleksei Aksimentiev. "De novo reconstruction of DNA origami structures through atomistic molecular dynamics simulation". In: *Nucleic Acids Res.* 44.7 (Apr. 2016), pp. 3013–3019.

[32] Massimo Kube et al. "Revealing the structures of megadalton-scale DNA complexes with nucleotide resolution". In: *Nat. Commun.* 11.1 (Dec. 2020), p. 6229.

[33] Xiao Chen Bai et al. "Cryo-EM structure of a 3D DNA-origami object". In: *Proc. Natl. Acad. Sci. U. S. A.* 109.49 (Dec. 2012), pp. 20012–20017.

[34] Vincent B. Chen et al. "MolProbity: All-atom structure validation for macromolecular crystallography". In: *Acta Crystallogr. Sect. D Biol. Crystallogr.* 66.1 (2010), pp. 12–21.

[35] Randy J. Read et al. "A new generation of crystallographic validation tools for the Protein Data Bank". In: *Structure* 19.10 (2011), pp. 1395–1412.

[36] Paul Emsley and Kevin Cowtan. "Coot : model-building tools for molecular graphics". In: *Acta Crystallogr. Sect. D Biol. Crystallogr.* 60.12 (Dec. 2004), pp. 2126–2132.

[37] Yuanchen Dong et al. "Folding DNA into a Lipid-Conjugated Nanobarrel for Controlled Reconstitution of Membrane Proteins". In: *Angew. Chemie - Int. Ed.* 57.8 (Feb. 2018), pp. 2072–2076.

[38] Jonathan P.K. Doye et al. "Coarse-graining DNA for simulations of DNA nanotechnology". In: *Phys. Chem. Chem. Phys.* 15.47 (2013), pp. 20395–20414.

[39] Jae Young Lee et al. "Rapid Computational Analysis of DNA Origami Assemblies at Near-Atomic Resolution". In: *ACS Nano* 15.1 (Jan. 2021), pp. 1002–1015.

[40] R. E. Franklin and R. G. Gosling. "The structure of sodium thymonucleate fibres. I. The influence of water content". In: *Acta Crystallogr.* 6.8 (Sept. 1953), pp. 673–677.

[41] M. H. F. WILKINS, A. R. STOKES, and H. R. WILSON. "Molecular Structure of Nucleic Acids: Molecular Structure of Deoxypentose Nucleic Acids". In: *Nature* 171.4356 (Apr. 1953), pp. 738–740.

[42] R. Langridge et al. "MOLECULAR STRUCTURE OF DEOXYRIBONUCLEIC ACID (DNA)". In: *J. Biophys. Biochem. Cytol.* 3.5 (Sept. 1957), pp. 767–778.

[43]   R. Langridge et al. "The molecular configuration of deoxyribonucleic acid". In: *J. Mol. Biol.* 2.1 (Apr. 1960), 19–IN11.

[44]   F. H. C. CRICK and J. D. WATSON. "The complementary structure of deoxyribonucleic acid". In: *Proc. R. Soc. London. Ser. A. Math. Phys. Sci.* 223.1152 (Apr. 1954), pp. 80–96.

[45]   A.G.W. Leslie et al. "Polymorphism of DNA double helices". In: *J. Mol. Biol.* 143.1 (Oct. 1980), pp. 49–72.

[46]   Matthew Meselson and Franklin W. Stahl. "The replication of DNA in Escherichia coli". In: *Proc. Natl. Acad. Sci.* 44.7 (July 1958), pp. 671–682.

[47]   J. Craig Venter et al. "The Sequence of the Human Genome". In: *Science (80-.).* 291.5507 (Feb. 2001), pp. 1304–1351.

[48]   Nadrian C. Seeman. "Nucleic acid junctions and lattices". In: *J. Theor. Biol.* 99.2 (Nov. 1982), pp. 237–247.

[49]   H. R. Drew et al. "Structure of a B-DNA dodecamer: conformation and dynamics." In: *Proc. Natl. Acad. Sci.* 78.4 (Apr. 1981), pp. 2179–2183.

[50]   C.R. Calladine and H.R. Drew. "Principles of sequence-dependent flexure of DNA". In: *J. Mol. Biol.* 192.4 (Dec. 1986), pp. 907–918.

[51]   Wolfram Saenger. "Structure and Dynamics of Water Surrounding Biomolecules". In: *Annu. Rev. Biophys. Biophys. Chem.* 16.1 (June 1987), pp. 93–114.

[52]   P. Shing and Megan Carter. "DNA Structure: Alphabet Soup for the Cellular Soul". In: *DNA Replication-Current Adv.* February (2011).

[53]   Darrin M. York et al. "Toward the Accurate Modeling of DNA: The Importance of Long-Range Electrostatics". In: *J. Am. Chem. Soc.* 117.17 (1995), pp. 5001–5002.

[54]   Wendy D. Cornell et al. "A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules". In: *J. Am. Chem. Soc.* 117.19 (May 1995), pp. 5179–5197.

[55]   Nicolas Foloppe and Alexander D. MacKerell, Jr. "All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data". In: *J. Comput. Chem.* 21.2 (Jan. 2000), pp. 86–104.

[56]   Thereza A. Soares et al. "An improved nucleic acid parameter set for the GROMOS force field". In: *J. Comput. Chem.* 26.7 (May 2005), pp. 725–737.

[57]   Matthew A. Young, G. Ravishanker, and D. L. Beveridge. "A 5-nanosecond molecular dynamics trajectory for B-DNA: Analysis of structure, motions, and solvation". In: *Biophys. J.* 73.5 (Nov. 1997), pp. 2313–2336.

[58]   Péter Várnai et al. *α/γ transitions in the B-DNA backbone*. Dec. 2002.

[59]   John SantaLucia and Donald Hicks. "The thermodynamics of DNA structural motifs". In: *Annu. Rev. Biophys. Biomol. Struct.* 33.1 (June 2004), pp. 415–440.

[60]   Pavel Hobza and Jiří Šponer. "Structure, Energetics, and Dynamics of the Nucleic Acid Base Pairs: Nonempirical Ab Initio Calculations". In: *Chem. Rev.* 99.11 (Nov. 1999), pp. 3247–3276.

[61]   X.-J. Lu. "3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures". In: *Nucleic Acids Res.* 31.17 (Sept. 2003), pp. 5108–5121.

[62]   Piotr Cysewski. "The post-SCF quantum chemistry characteristics of inter- and intra-strand stacking interactions in d(CpG) and d(GpC) steps found in B-DNA, A-DNA and Z-DNA crystals". In: *J. Mol. Model.* 15.6 (June 2009), pp. 597–606.

[63]   Jiří Šponer, Kevin E. Riley, and Pavel Hobza. "Nature and magnitude of aromatic stacking of nucleic acid bases". In: *Phys. Chem. Chem. Phys.* 10.19 (2008), p. 2595.

[64]   R.E. Dickerson. "Definitions and nomenclature of nucleic acid structure components". In: *Nucleic Acids Res.* 17.5 (1989), pp. 1797–1803.

[65]   Wilma K. Olson et al. "A standard reference frame for the description of nucleic acid base-pair geometry". In: *J. Mol. Biol.* 313.1 (Oct. 2001), pp. 229–237.

[66]   Andrey A. Gorin, Victor B. Zhurkin, and Wilma K. "B-DNA Twisting Correlates with Base-pair Morphology". In: *J. Mol. Biol.* 247.1 (Mar. 1995), pp. 34–48.

[67]   Pablo D. Dans et al. "The static and dynamic structural heterogeneities of B-DNA: extending Calladine–Dickerson rules". In: *Nucleic Acids Res.* 47.21 (Dec. 2019), pp. 11090–11102.

[68]   Marco Pasi et al. "μABC: a systematic microsecond molecular dynamics study of tetranucleotide sequence effects in B-DNA". In: *Nucleic Acids Res.* 42.19 (Oct. 2014), pp. 12272–12283.

[69]   Gabriela da Rosa et al. "Sequence-dependent structural properties of B-DNA: what have we learned in 40 years?" In: *Biophys. Rev.* 13.6 (Dec. 2021), pp. 995–1005.

[70]   Korbinian Liebl and Martin Zacharias. "Accurate modeling of DNA conformational flexibility by a multivariate Ising model". In: *Proc. Natl. Acad. Sci.* 118.15 (Apr. 2021), pp. 1–8.

[71]   Jack W. Shepherd et al. "The emergence of sequence-dependent structural motifs in stretched, torsionally constrained DNA". In: *Nucleic Acids Res.* 48.4 (Feb. 2020), pp. 1748–1763.

[72]   Alexandra Balaceanu et al. "Modulation of the helical properties of DNA: next-to-nearest neighbour effects and beyond". In: *Nucleic Acids Res.* 47.9 (May 2019), pp. 4418–4430.

[73]   David L. Beveridge, Thomas E. Cheatham, and Mihaly Mezei. "The ABCs of molecular dynamics simulations on B-DNA, circa 2012". In: *J. Biosci.* 37.3 (July 2012), pp. 379–397.

[74]   Jiří Šponer et al. "The DNA and RNA sugar-phosphate backbone emerges as the key player. An overview of quantum-chemical, structural biology and simulation studies". In: *Phys. Chem. Chem. Phys.* 14.44 (2012), pp. 15257–15277.

[75]   Katarina Hart et al. "Optimization of the CHARMM Additive Force Field for DNA: Improved Treatment of the BI/BII Conformational Equilibrium". In: *J. Chem. Theory Comput.* 8.1 (Jan. 2012), pp. 348–362.

[76]   Benedict E.K. Snodin et al. "Introducing improved structural properties and salt dependence into a coarse-grained model of DNA". In: *J. Chem. Phys.* 142.23 (June 2015), p. 234901.

[77]   Robin Holliday. "A mechanism for gene conversion in fungi". In: *Genet. Res.* 5.2 (July 1964), pp. 282–304.

[78]   Junghuei Chen and Nadrian C Seeman. "Synthesis from DNA of a molecule with the connectivity of a cube". In: *Nature* 350.6319 (Apr. 1991), pp. 631–633.

[79]   Fabian Kohler. "Precision Design and Characterization of DNA Origami Nanostructures Using High-Resolution Cryo-EM". PhD thesis. 2021.

[80]   Brandt F. Eichman et al. "The Holliday junction in an inverted repeat DNA sequence: Sequence effects on the structure of four-way junctions". In: *Proc. Natl. Acad. Sci. U. S. A.* 97.8 (2000), pp. 3971–3976.

[81]   Yonggang Ke et al. "Multilayer DNA origami packed on a square lattice". In: *J. Am. Chem. Soc.* 131.43 (2009), pp. 15903–15908.

[82]   Yonggang Ke et al. "Multilayer DNA origami packed on hexagonal and hybrid lattices". In: *J. Am. Chem. Soc.* 134.3 (2012), pp. 1770–1774.

[83]   Dongran Han et al. "DNA Gridiron Nanostructures Based on Four-Arm Junctions". In: *Science (80-.).* 339.6126 (Mar. 2013), pp. 1412–1415.

[84]   Wen Wang et al. "Complex wireframe DNA nanostructures from simple building blocks". In: *Nat. Commun.* 10.1 (Dec. 2019), p. 1067.

[85]   Thomas G. Martin and Hendrik Dietz. "Magnesium-free self-assembly of multi-layer DNA objects". In: *Nat. Commun.* 3 (2012).

[86]   Klaus F. Wagenbauer et al. "How We Make DNA Origami". In: *ChemBioChem* 18.19 (Oct. 2017), pp. 1873–1885.

[87]   Jean-Philippe J. Sobczak et al. "Rapid Folding of DNA into Nanoscale Shapes at Constant Temperature". In: *Science (80-.).* 338.6113 (Dec. 2012), pp. 1458–1461.

[88]   Fabian Schneider, Natalie Möritz, and Hendrik Dietz. "The sequence of events during folding of a DNA origami". In: *Sci. Adv.* 5.5 (2019), eaaw1412.

[89]   Nayan P. Agarwal et al. "Block Copolymer Micellization as a Protection Strategy for DNA Origami". In: *Angew. Chemie Int. Ed.* 56.20 (May 2017), pp. 5460–5464.

[90]   Florian Praetorius et al. "Biotechnological mass production of DNA origami". In: *Nature* 552.7683 (2017), pp. 84–87.

[91]   Floris A.S. Engelhardt et al. "Custom-Size, Functional, and Durable DNA Origami with Design-Specific Scaffolds". In: *ACS Nano* 13.5 (2019), pp. 5015–5027.

[92]   Benjamin Kick et al. "Efficient Production of Single-Stranded Phage DNA as Scaffolds for DNA Origami". In: *Nano Lett.* 15.7 (July 2015), pp. 4672–4676.

[93]   Philipp C Nickels et al. "Molecular force spectroscopy with a DNA origami–based nanoscopic force clamp". In: *Science (80-.).* 354.6310 (Oct. 2016), pp. 305–307.

[94]   Joachim Frank. *Three-dimensional electron microscopy of macromolecular assemblies: visualization of biological molecules in their native state.* Oxford university press, 2006.

[95]   Sjors H.W. Scheres. "RELION: Implementation of a Bayesian approach to cryo-EM structure determination". In: *J. Struct. Biol.* 180.3 (Dec. 2012), pp. 519–530.

[96]   Sjors H.W. Scheres. "A Bayesian View on Cryo-EM Structure Determination". In: *J. Mol. Biol.* 415.2 (Jan. 2012), pp. 406–418.

[97]   Dari Kimanius et al. "Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2". In: *Elife* 5 (Nov. 2016), pp. 1–21.

[98]   Jasenko Zivanov et al. "New tools for automated high-resolution cryo-EM structure determination in RELION-3". In: *Elife* 7 (Nov. 2018), pp. 1–22.

[99]   Takanori Nakane et al. "Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION". In: *Elife* 7 (2018), pp. 1–18.

[100]   Takanori Nakane et al. "Single-particle cryo-EM at atomic resolution". In: *Nature* 587.7832 (Nov. 2020), pp. 152–156.

[101]   David Bhella. "Cryo-electron microscopy: an introduction to the technique, and considerations when working to establish a national facility". In: *Biophys. Rev.* 11.4 (Aug. 2019), pp. 515–519.

[102]   RN Bracewell. "Strip Integration in Radio Astronomy". In: *Aust. J. Phys.* 9.2 (1956), p. 198.

[103]   Chad R. Simmons et al. "A Self-Assembled Rhombohedral DNA Crystal Scaffold with Tunable Cavity Sizes and High-Resolution Structural Detail". In: *Angew. Chemie* 132.42 (Oct. 2020), pp. 18778–18785.

[104]   Aneesur Rahman. "Correlations in the Motion of Atoms in Liquid Argon". In: *Phys. Rev.* 136.2A (Oct. 1964), A405–A411.

[105]   Dennis C Rapaport. *The art of molecular dynamics simulation.* Cambridge university press, 2004.

[106]   Daan Frenkel and Berend Smit. "Understanding molecular simulation: From algorithms to applications". In: *Underst. Mol. Simul. From algorithms to Appl.* 1 (1996).

[107]   W. G. Hoover, H. A. Posch, and S. Bestiale. "Dense-fluid Lyapunov spectra via constrained molecular dynamics". In: *J. Chem. Phys.* 87.11 (Dec. 1987), pp. 6665–6670.

[108]   D. Levesque and L. Verlet. "Molecular dynamics and time reversibility". In: *J. Stat. Phys.* 72.3-4 (Aug. 1993), pp. 519–537.

[109]   William C. Swope et al. "A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters". In: *J. Chem. Phys.* 76.1 (Jan. 1982), pp. 637–649.

[110]   Loup Verlet. "Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules". In: *Phys. Rev.* 159.1 (July 1967), pp. 98–103.

[111]   Jones J.E. "On the determination of molecular fields.—I. From the variation of the viscosity of a gas with temperature". In: *Proc. R. Soc. London. Ser. A, Contain. Pap. a Math. Phys. Character* 106.738 (Oct. 1924), pp. 441–462.

[112]   J.E. Lennard-Jones. "Cohesion". In: *Proc. Phys. Soc.* 43.461 (1931).

[113]   Scott J. Weiner et al. "An all atom force field for simulations of proteins and nucleic acids". In: *J. Comput. Chem.* 7.2 (Apr. 1986), pp. 230–252.

[114]   Pablo D. Dans et al. "How accurate are accurate force-fields for B-DNA?" In: *Nucleic Acids Res.* 45.7 (2017), pp. 4217–4230.

[115]   Rodrigo Galindo-Murillo et al. "Assessing the Current State of Amber Force Field Modifications for DNA". In: *J. Chem. Theory Comput.* 12.8 (Aug. 2016), pp. 4114–4127.

[116] Christopher Maffeo and Aleksei Aksimentiev. "MrDNA: A multi-resolution model for predicting the structure and dynamics of DNA systems". In: *Nucleic Acids Res.* 48.9 (May 2020), pp. 5135–5146.

[117] A. Sengar et al. "A Primer on the oxDNA Model of DNA: When to Use it, How to Simulate it and How to Interpret the Results". In: *Front. Mol. Biosci.* 8.June (June 2021), pp. 1–22.

[118] Thomas E. Ouldridge, Ard A. Louis, and Jonathan P. K. Doye. "Structural, mechanical and thermodynamic properties of a coarse-grained DNA model". In: *J. Am. Chem. Soc.* 117.19 (Sept. 2010), pp. 5179–5197.

[119] Petr Šulc et al. "Sequence-dependent thermodynamics of a coarse-grained DNA model". In: *J. Chem. Phys.* 137.13 (Oct. 2012), p. 135101.

[120] Lorenzo Rovigatti et al. "A comparison between parallelization approaches in molecular dynamics simulations on GPUs". In: *J. Comput. Chem.* 36.1 (Jan. 2015), pp. 1–8.

[121] Antonio Suma et al. "TacoxDNA: A user-friendly web server for simulations of complex DNA structures, from single strands to origami". In: *J. Comput. Chem.* 40.29 (Nov. 2019), pp. 2586–2595.

[122] Joakim Bohlin et al. "Design and simulation of DNA, RNA and hybrid protein–nucleic acid nanostructures with oxView". In: *Nat. Protoc.* (June 2022), pp. 1–27.

[123] Petr Šulc et al. "A nucleotide-level coarse-grained model of RNA". In: *J. Chem. Phys.* 140.23 (June 2014), p. 235102.

[124] Christian Matek et al. "Coarse-grained modelling of supercoiled RNA". In: *J. Chem. Phys.* 143.24 (Dec. 2015), p. 243122.

[125] Jeffrey Comer and Aleksei Aksimentiev. "Predicting the DNA Sequence Dependence of Nanopore Ion Current Using Atomic-Resolution Brownian Dynamics". In: *J. Phys. Chem. C* 116.5 (Feb. 2012), pp. 3376–3393.

[126] Do Nyun Kim et al. "Quantitative prediction of 3D solution shape and flexibility of nucleic acid nanostructures". In: *Nucleic Acids Res.* 40.7 (Apr. 2012), pp. 2862–2868.

[127] Keyao Pan et al. "Structure and conformational dynamics of scaffolded DNA origami nanoparticles". In: *Nucleic Acids Res.* 45.11 (June 2017), pp. 6284–6298.

[128] Jae Gyung Lee et al. "Predicting the Free-Form Shape of Structured DNA Assemblies from Their Lattice-Based Design Blueprint". In: *ACS Nano* 16.3 (Mar. 2022), pp. 4289–4297.

[129] John W Tukey. *Exploratory Data Analysis.* Vol. 2. Reading, MA, 1977.

[130]   Leonardo G. Trabuco et al. "Flexible Fitting of Atomic Structures into Electron Microscopy Maps Using Molecular Dynamics". In: *Structure* 16.5 (May 2008), pp. 673–683.

[131]   Wayne A. Hendrickson. "Stereochemically restrained refinement of macromolecular structures". In: *Methods Enzymol.* Vol. 115. C. 1985, pp. 252–270.

[132]   Willy Wriggers, Ronald A Milligan, and J.Andrew McCammon. "Situs: A Package for Docking Crystal Structures into Low-Resolution Maps from Electron Microscopy". In: *J. Struct. Biol.* 125.2-3 (Apr. 1999), pp. 185–195.

[133]   Niels Volkmann and Dorit Hanein. "Quantitative Fitting of Atomic Models into Observed Densities Derived by Electron Microscopy". In: *J. Struct. Biol.* 125.2-3 (Apr. 1999), pp. 176–184.

[134]   Michael G. Rossmann. "Fitting atomic models into electron-microscopy maps". In: *Acta Crystallogr. Sect. D Biol. Crystallogr.* 56.10 (Oct. 2000), pp. 1341–1349.

[135]   Wen Jiang et al. "Bridging the information gap: Computational tools for intermediate resolution structure interpretation". In: *J. Mol. Biol.* 308.5 (2001), pp. 1033–1044.

[136]   Michael G. Rossmann, Ricardo Bernal, and Sergei V. Pletnev. "Combining Electron Microscopic with X-Ray Crystallographic Structures". In: *J. Struct. Biol.* 136.3 (Dec. 2001), pp. 190–200.

[137]   Pablo Chacón and Willy Wriggers. "Multi-resolution contour-based fitting of macromolecular structures". In: *J. Mol. Biol.* 317.3 (Mar. 2002), pp. 375–384.

[138]   Ryan Mcgreevy et al. "XMDFF: Molecular dynamics flexible fitting of low-resolution X-ray structures". In: *Acta Crystallogr. Sect. D Biol. Crystallogr.* 70.9 (2014), pp. 2344–2355.

[139]   James Z. Chen et al. "Low-resolution structure refinement in electron microscopy". In: *J. Struct. Biol.* 144.1-2 (Oct. 2003), pp. 144–151.

[140]   Gunnar F. Schröder, Axel T. Brunger, and Michael Levitt. "Combining Efficient Conformational Sampling with a Deformable Elastic Network Model Facilitates Structure Refinement at Low Resolution". In: *Structure* 15.12 (Dec. 2007), pp. 1630–1641.

[141]   Willy Wriggers and Pablo Chacón. "Modeling tricks and fitting techniques for multiresolution structures". In: *Structure* 9.9 (Sept. 2001), pp. 779–788.

[142]   Willy Wriggers and Stefan Birmanns. "Using Situs for Flexible and Rigid-Body Fitting of Multiresolution Single-Molecule Data". In: *J. Struct. Biol.* 133.2-3 (Feb. 2001), pp. 193–202.

[143]   Marek Orzechowski and Florence Tama. "Flexible fitting of high-resolution x-ray structures into cryoelectron microscopy maps using biased molecular dynamics simulations". In: *Biophys. J.* 95.12 (2008), pp. 5692–5705.

[144]   B. Isralewitz, M. Gao, and K. Schulten. "Steered molecular dynamics and mechanical functions of proteins Isralewitz, Gao and Schulten 225". In: *Curr. Opin. Struct. Biol.* 11 (2001), pp. 224–230.

[145]   Leonardo G. Trabuco et al. "Molecular dynamics flexible fitting: A practical guide to combine cryo-electron microscopy and X-ray crystallography". In: *Methods* 49.2 (Oct. 2009), pp. 174–180.

[146]   John E. Stone et al. "GPU-accelerated analysis and visualization of large structures solved by molecular dynamics flexible fitting". In: *Faraday Discuss.* 169 (2014), pp. 265–283.

[147]   Abhishek Singharoy et al. "Molecular dynamics-based refinement and validation for sub-5 Å cryo-electron microscopy maps". In: *Elife* 5.JULY (July 2016), pp. 1–32.

[148]   Osamu Miyashita et al. "Flexible fitting to cryo-EM density map using ensemble molecular dynamics simulations". In: *J. Comput. Chem.* 38.16 (2017), pp. 1447–1461.

[149]   Maxim Igaev et al. "Automated cryo-EM structure refinement using correlation-driven molecular dynamics". In: *Elife* 8 (Mar. 2019), pp. 1–33.

[150]   Marcos Sotomayor and Klaus Schulten. "Single-Molecule Experiments in Vitro and in Silico". In: *Science (80-.).* 316.5828 (May 2007), pp. 1144–1148.

[151]   David B. Wells, Volha Abramkina, and Aleksei Aksimentiev. "Exploring transmembrane transport through α -hemolysin with grid-steered molecular dynamics". In: *J. Chem. Phys.* 127.12 (Sept. 2007), p. 125101.

[152]   Elizabeth Villa et al. "Ribosome-induced changes in elongation factor Tu conformation control GTP hydrolysis". In: *Proc. Natl. Acad. Sci.* 106.4 (Jan. 2009), pp. 1063–1068.

[153]   Florence Tama, Osamu Miyashita, and Charles L. Brooks III. "Normal mode based flexible fitting of high-resolution structure into low-resolution experimental data from cryo-EM". In: *J. Struct. Biol.* 147.3 (Sept. 2004), pp. 315–326.

[154]   P.L. Stewart, S.D. Fuller, and R.M. Burnett. "Difference imaging of adenovirus: bridging the resolution gap between X-ray crystallography and electron microscopy." In: *EMBO J.* 12.7 (July 1993), pp. 2589–2599.

[155] Alan M. Roseman. "Docking structures of domains into maps from cryo-electron microscopy using local correlation". In: *Acta Crystallogr. Sect. D Biol. Crystallogr.* 56.10 (Oct. 2000), pp. 1332–1340.

[156] Jian Sheng Jiang and Axel T. Brünger. *Protein hydration observed by X-ray diffraction: Solvation properties of penicillopepsin and neuraminidase crystal structures.* 1994.

[157] Pavel V. Afonine et al. "New tools for the analysis and validation of cryo-EM maps and atomic models". In: *Acta Crystallogr. Sect. D Struct. Biol.* 74.9 (2018), pp. 814–840.

[158] H. M. Berman. "The Protein Data Bank". In: *Nucleic Acids Res.* 28.1 (Jan. 2000), pp. 235–242.

[159] Yuhang Wang et al. "Constructing atomic structural models into cryo-EM densities using molecular dynamics – Pros and cons". In: *J. Struct. Biol.* 204.2 (Nov. 2018), pp. 319–328.

[160] F. DiMaio and W. Chiu. "Tools for Model Building and Optimization into Near-Atomic Resolution Electron Cryo-Microscopy Density Maps". In: *Methods Enzymol.* 579.1 (2016), pp. 255–276.

[161] Martyn D. Winn et al. "Overview of the CCP 4 suite and current developments". In: *Acta Crystallogr. Sect. D Biol. Crystallogr.* 67.4 (Apr. 2011), pp. 235–242.

[162] Igor Orlov et al. "The integrative role of cryo electron microscopy in molecular and cellular structural biology". In: *Biol. Cell* 109.2 (Feb. 2017), pp. 81–93.

[163] Pavel V. Afonine et al. "Real-space refinement in PHENIX for cryo-EM and crystallography". In: *Acta Crystallogr. Sect. D Struct. Biol.* 74.6 (June 2018), pp. 531–544.

[164] Arie Kaufman et al. "Research issues in Volume Visualization". In: *IEEE Comput. Graph. Appl.* 14.2 (Mar. 1994), pp. 63–67.

[165] James C. Phillips et al. "Scalable molecular dynamics with NAMD". In: *Parallel Sci. Eng. Appl. Charm++ Approach.* Vol. 26. 16. CRC Press, Apr. 2016, pp. 60–76.

[166] William Humphrey, Andrew Dalke, and Klaus Schulten. "VMD: Visual molecular dynamics". In: *J. Mol. Graph.* 14.1 (1996), pp. 33–38.

[167] Roman V. Reshetnikov et al. "A coarse-grained model for DNA origami". In: *Nucleic Acids Res.* 46.3 (Feb. 2018), pp. 1102–1112.

[168] Peter B. Rosenthal and John L. Rubinstein. "Validating maps from single particle electron cryomicroscopy". In: *Curr. Opin. Struct. Biol.* 34 (2015), pp. 135–144.

[169] Philip E. Bourne et al. "[30] Macromolecular crystallographic information file". In: *Methods Enzymol.* Vol. 277. 1997, pp. 571–590.

[170] Daichi Hayakawa et al. "Geometrically programmed self-limited assembly of tubules using DNA origami colloids". In: *ArXiv:* 2203.01421 (2022).

[171] Bireswar Laha and Doug A Bowman. "Identifying the Benefits of Immersion in Virtual Reality for Volume Data Visualization". In: *Immersive Vis. Revisit. Work. IEEE VR Conf.* D.March (2012), pp. 1–2.

[172] Mohamed El Beheiry et al. "Virtual Reality: Beyond Visualization". In: *J. Mol. Biol.* 431.7 (Mar. 2019), pp. 1315–1321.

[173] Thomas Kluyver et al. "Jupyter Notebooks—a publishing format for reproducible computational workflows". In: *Position. Power Acad. Publ. Play. Agents Agendas - Proc. 20th Int. Conf. Electron. Publ. ELPUB 2016* (2016), pp. 87–90.

[174] Hugues Hoppe et al. "Surface reconstruction from unorganized points". In: *Proc. 19th Annu. Conf. Comput. Graph. Interact. Tech. - SIGGRAPH '92.* Vol. 4553. July 1992. New York, New York, USA: ACM Press, 1992, pp. 71–78.

[175] Eric F. Pettersen et al. "UCSF Chimera - A visualization system for exploratory research and analysis". In: *J. Comput. Chem.* 25.13 (2004), pp. 1605–1612.

[176] Naveen Michaud-Agrawal et al. "MDAnalysis: A toolkit for the analysis of molecular dynamics simulations". In: *J. Comput. Chem.* 32.10 (July 2011), pp. 2319–2327.

[177] Tom Burnley, Colin M. Palmer, and Martyn Winn. "Recent developments in the CCP-EM software suite". In: *Acta Crystallogr. Sect. D Struct. Biol.* 73.6 (June 2017), pp. 469–477.

[178] John D. Hunter. "Matplotlib: A 2D Graphics Environment". In: *Comput. Sci. Eng.* 9.3 (2007), pp. 90–95.

[179] Charlotte Kielar et al. "Dynamics of DNA Origami Lattice Formation at Solid-Liquid Interfaces". In: *ACS Appl. Mater. Interfaces* 10.51 (Dec. 2018), pp. 44844–44853.

[180] Paul S. Kwon et al. "Designer DNA architecture offers precise and multivalent spatial pattern-recognition for viral sensing and inhibition". In: *Nat. Chem.* 12.1 (Jan. 2020), pp. 26–35.

[181] Veikko Linko, Marika Eerikäinen, and Mauri A. Kostiainen. "A modular DNA origami-based enzyme cascade nanoreactor". In: *Chem. Commun.* 51.25 (2015), pp. 5351–5354.

[182] Jian-sheng Sun, Thérèse Carestier, and Claude Hélène. "Oligonucleotide directed triple helix formation". In: *Curr. Opin. Struct. Biol.* 6.3 (June 1996), pp. 327–333.

[183]   Sangkee Rhee et al. "Structure of a triple helical DNA with a triplex-duplex junction". In: *Biochemistry* 38.51 (1999), pp. 16810–16815.

[184]   K Hoogsteen. "The structure of crystals containing a hydrogen-bonded complex of 1-methylthymine and 9-methyladenine". In: *Acta Crystallogr.* 12.10 (Oct. 1959), pp. 822–823.

[185]   Satoshi Fujii et al. "Sequence-dependent DNA deformability studied using molecular dynamics simulations". In: *Nucleic Acids Res.* 35.18 (2007), pp. 6063–6074.

[186]   Richard E. Dickerson and A. Klug. "Base sequence and helix structure variation in B and A DNA". In: *J. Mol. Biol.* 166.3 (May 1983), pp. 419–441.

[187]   Wilma K. Olson et al. "DNA sequence-dependent deformability deduced from protein–DNA crystal complexes". In: *Proc. Natl. Acad. Sci.* 95.19 (Sept. 1998), pp. 11163–11168.

[188]   Filip Lankaš et al. "DNA Deformability at the Base Pair Level". In: *J. Am. Chem. Soc.* 126.13 (Apr. 2004), pp. 4124–4125.

[189]   Marie Zgarbová et al. "Toward Improved Description of DNA Backbone: Revisiting Epsilon and Zeta Torsion Force Field Parameters". In: *J. Chem. Theory Comput.* 9.5 (May 2013), pp. 2339–2354.

[190]   Tomáš Dršata and Filip Lankaš. "Theoretical models of DNA flexibility". In: *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 3.4 (July 2013), pp. 355–363.

[191]   Jing Huang et al. "CHARMM36m: An improved force field for folded and intrinsically disordered proteins". In: *Nat. Methods* 14.1 (Jan. 2016), pp. 71–73.

[192]   Korbinian Liebl and Martin Zacharias. "How methyl–sugar interactions determine DNA structure and flexibility". In: *Nucleic Acids Res.* 47.3 (Feb. 2019), pp. 1132–1140.

[193]   Alberto Pérez, F. Javier Luque, and Modesto Orozco. "Frontiers in Molecular Dynamics Simulations of DNA". In: *Acc. Chem. Res.* 45.2 (Feb. 2012), pp. 196–205.

[194]   Benedict E K Snodin et al. "Coarse-grained modelling of the structural properties of DNA origami". In: *Nucleic Acids Res.* 47.3 (Feb. 2019), pp. 1585–1597.

[195]   Swarup Dey et al. "DNA origami". In: *Nat. Rev. Methods Prim.* 1.1 (2021), pp. 1–38.

[196]   Martin Glaser et al. "Review the art of designing dna nanostructures with cad software". In: *Molecules* 26.8 (2021), pp. 1–24.

[197]   Jejoong Yoo and Aleksei Aksimentiev. "In situ structure and dynamics of DNA origami determined through molecular dynamics simulations". In: *Proc. Natl. Acad. Sci. U. S. A.* 110.50 (2013), pp. 20099–20104.

[198] Jejoong Yoo and Aleksei Aksimentiev. "Molecular Dynamics of Membrane-Spanning DNA Channels: Conductance Mechanism, Electro-Osmotic Transport, and Mechanical Gating". In: *J. Phys. Chem. Lett.* 6.23 (2015), pp. 4680–4687.

[199] Hamid Ramezani and Hendrik Dietz. "Building machines with DNA molecules". In: *Nat. Rev. Genet.* 21.1 (2020), pp. 5–26.

[200] Thomas Dodd, Chunli Yan, and Ivaylo Ivanov. "Simulation-Based Methods for Model Building and Refinement in Cryoelectron Microscopy". In: *J. Chem. Inf. Model.* 60.5 (May 2020), pp. 2470–2483.

[201] Joana Pereira et al. "High-accuracy protein structure prediction in CASP14". In: *Proteins Struct. Funct. Bioinforma.* 89.12 (Dec. 2021), pp. 1687–1699.

[202] John Jumper et al. "Highly accurate protein structure prediction with AlphaFold". In: *Nature* 596.7873 (2021), pp. 583–589.

[203] Philipp Mostosi et al. "Haruspex: A Neural Network for the Automatic Identification of Oligonucleotides and Protein Secondary Structure in Cryo-Electron Microscopy Maps". In: *Angew. Chemie - Int. Ed.* 59.35 (2020), pp. 14788–14795.

[204] K. Vanommeslaeghe et al. "CHARMM.pdf". In: *J. Comput. Chem.* 31.4 (2010), pp. 671–690.

[205] Eric F. Pettersen et al. "UCSF ChimeraX: Structure visualization for researchers, educators, and developers". In: *Protein Sci.* 30.1 (Jan. 2021), pp. 70–82.

Acknowledgment

My gratitude goes to everyone who supported and encouraged me. My achievements are as much your attire as yours feel like mine.

Hendrik, for your conviction that simulations are not just fiction, but provide extensive insight and support for experimental processes, hence including a computational physicist into a team of skillful experimentalists. For providing excellent equipment and numerous opportunities for contributing to a large variety of projects within the group. For the great support, motivating advice, and repeated affirmations of how much you appreciate my contributions to the group.

The Max Planck School Matter to Life, for the financial support, but more so for a chance to work collaboratively on a bigger goal with like-minded students from all over the country.

Massimo and Fabian, for cooking up delight in the FrozenGulasch efforts, providing exceptional experimental data to show off to friends and family. Max, for experiencing the joy of collaborative synergy between experimentalist and theoretician with you. We both know that experimentalists go first, but most computational scientists are zero-indexed. Max, Fabian, and Chris, for letting me draft in your slipstream when cycling to work. All dear colleagues and friends from DietzLab, for joint adventures in DNA nanotechnology and other activities in and around Munich.

Eva, Elena, Marlene, Fabian, Max, and Tobi, for proofreading this manuscript - with Tobi and Marlene completing the proofreading triple.

In admiration to my loving parents and brothers for always providing support and refuge, and being my closest friends. Finally, to those who are my family despite not sharing the same parents. Thank you.