

The exclusive license for this PDF is limited to personal website use only. No part of this digital document may be reproduced, stored in a retrieval system or transmitted commercially in any form or by any means. The publisher has taken reasonable care in the preparation of this digital document, but makes no expressed or implied warranty of any kind and assumes no responsibility for any errors or omissions. No liability is assumed for incidental or consequential damages in connection with or arising out of information contained herein. This digital document is sold with the clear understanding that the publisher is not engaged in rendering legal, medical or any other professional services.

Chapter 8

UTILIZATION OF COMPUTER SCIENCE FOR CONSTRUCTION AND CHARACTERIZATION OF DNA NANO-STRUCTURES

*Felicie F. Andersen^{#1}, Palle Villesen^{#2}, Bjarne Knudsen³,
Carsten Wiuf², Alessandro Desideri⁴, Birgitta R. Knudsen^{1*}*

¹Department of Molecular Biology and Interdisciplinary Nanoscience Center (iNANO), Aarhus University, C.F. Moellers Alle 3, Bldg. 1130, DK-8000 Aarhus C, Denmark

²Bioinformatics Research Centre, Aarhus University, C.F. Moellers Alle 8, Bldg. 1110, DK-8000 Aarhus C, Denmark.

³CLC bio A/S, Finlandsgade 10-12, DK-8200 Aarhus N, Denmark

⁴Department of Biology and NAST Nanoscience and Nanotechnology and Innovative Instrumentation, University of Rome "Tor Vergata", Via della Ricerca Scientifica 1, 00133 Rome, Italy

ABSTRACT

In the upcoming field of DNA nano-science, DNA is used as an entity for building higher order self-assembled structures rather than for storage of genetic information. The application of DNA as a building block relies, like its biological blueprinting function, on the specific pairing between bases holding pairs of DNA molecules together. However, the complexity involved in building even simple DNA architectures is usually so high that designing structures from more than a few base sequences makes the use of computer programs indispensable. Moreover, synthetic DNA nano-structures are often time-consuming and expensive to build and their structures can be difficult to validate. Therefore, atomistic simulations, which can predict assembly efficiency and physical properties of the designed structures are of great value for both the processes of design and characterization. In this article we summarize examples of computational tools for

[#]FFA and PV contributed equally to this work.

^{*} Corresponding authors. Contact information AD: e-mail: desideri@uniroma2.it, Telephone: +39-06-72594376, Fax: +39-06-2022798; BRK: e-mail: brk@mb.au.dk, Telephone: +45-89422608, Fax: +45-89422612.

the design and characterization of DNA nano-structures with particular emphasis on a three-dimensional truncated octahedral structure.

INTRODUCTION

During recent years the number of presented two- (2D) or three-dimensional (3D) DNA nano-structures has exploded. The field of DNA nano-science was initiated in the early 1990's by Seeman, who presented the design and construction of a DNA nano-structure with the connectivity of a cube (Chen and Seeman 1991). Since then a tremendous amount of DNA nano-structures including 2D origamis (Yan *et al.*, 2003, Rothmund, 2006, Andersen *et al.*, 2008a, Voigt *et al.*, 2010) and tiles (LaBean, *et al.*, 2000, Mathieu *et al.*, 2005) or 3D polyhedra (Andersen *et al.*, 2008b, Goodman *et al.*, 2005, Shih *et al.*, 2004, He *et al.*, 2008, Bhatia *et al.*, 2009, Aldaye and Sleiman., 2007), origami-based honeycomb or box shapes (Douglas *et al.*, 2009a, Andersen *et al.*, 2009) and crystals (Zheng *et al.*, 2009) have been published. All these structures are based either on reciprocal exchange of DNA strands exemplified by the PX motif used for origami folding (Yan *et al.*, 2003, Rothmund, 2006, Andersen *et al.*, 2008a, Voigt *et al.*, 2010) and by the Holiday junction motif of several 3D polyhedra (Chen and Seeman, 1991, Andersen *et al.*, 2008b, Goodman *et al.*, 2005) or on synthetic branching of helices exemplified by prism structures branched by amidite derivatives (Aldaye and Sleiman, 2007). Regardless the branching motif of the structures, they all rely on self-assembly directed by the specific base pairing (G-C or A-T pairs) of DNA. The key to successful construction is to design the DNA sequences not only so that they are able to form the intended product but also so that no other product will be competitive to the target. To achieve this goal it is necessary to estimate the thermodynamics of all possible pairings of DNA sequences and select the sequences so that the desired product is by far the most thermodynamically favourable one. Obviously, the complexity of doing so even when designing rather simple structures, involving more than a few DNA strands, by far exceeds the capacity of the human mind. Therefore, in most cases the design of DNA sequences for the construction of nano-structures relies on more or less sophisticated computational tools in order to rule out sequence combinations prone to form unwanted structures. For some purposes generic tools, such the NCBI blast program etc., may suffice. However, when many DNA sequences are to be matched to form more complicated structures such programs are inadequate. In line with this, custom-made software developed specifically with the design of a particular type of DNA structure in mind are increasing in number. Examples of such software are programs dedicated for design of 2D or 3D DNA origami structures (Andersen *et al.*, 2008a, Douglas *et al.*, 2009b) or for minimizing unwanted annealing of DNA strands in general (Deaton, *et al.*, 2003, Goodman, 2005).

Common for all DNA nano-structures presented until date is that they rely at least to some extent on synthetic DNA oligonucleotides, which makes their construction rather expensive. This fact, taken together with some of the common analysis techniques, such as Small Angle X-ray Scattering (SAXS) and Cryo-Transmission Electron Microscopy (Cryo-TEM) requiring quite large amounts of material (Oliviera *et al.*, 2010, Andersen *et al.*, 2008b, Shih *et al.*, 2004, Kato *et al.*, 2009, Andersen *et al.*, 2009) pose a serious challenge to the validation of the structures. Thus, to counter such obstacles atomistic simulations, which can predict the likelihood of successful assembly as well as structural properties of DNA nano-

structures before experiments, are of great value. This statement is supported by atomic-level simulations of some of Seeman's DNA nano-structures (Maiti *et al.*, 2006) and, as will be described in more detail below, of our truncated octahedral DNA cage (Oliviera *et al.*, 2010, Falconi *et al.*, 2009). Both sets of simulations are validated by experimental data. Moreover, by joining geometric modelling and sequence optimization Kumara *et al.*, succeeded in developing new and efficient assembly pathways for double crossover motifs (Kumara *et al.*, 2008). Pointing towards the future, the software package GIDEON allows 3D modelling and simulation of DNA nano-structures (Birac *et al.*, 2006).

COMPUTATIONAL TOOL (MIOS) FOR DESIGN OF DNA NANO-STRUCTURES

Structural designs: Here we describe the simple program MIOS (MInimizing Oligo Similarities) as an example to illustrate the basic requirements for programs designated for the design of 3D polyhedral DNA nano-structures. Based on the number and length of annealing sequences in a DNA nano-structure of a chosen 3D geometry, MIOS searches for sequence sets with minimal unwanted base pairing of two types: i) annealing between different sequences (and their reverse complementary counterparts) and ii) self-annealing (hairpin formation).

To demonstrate the capabilities of MIOS we have used the program to design DNA oligonucleotide sequences for one-step assembly of a tetrahedron, a cube and an octahedron. The basic design of these structures (see schematic folding models, figure 1) relies on the same structural framework as our previously published octahedral DNA nano-cage (Andersen, *et al.*, 2008b, Oliveira *et al.*, 2010, Falconi *et al.*, 2009), with edges consisting of 18 base pairs double-stranded DNA and truncated corners consisting of three-thymidine long single-stranded stretches. All the presented 3D structures are assembled from linear oligonucleotides designed in such a way that the ends of each oligonucleotide are brought together by the specific base pairing. This circularization allows covalent sealing of all DNA ends upon assembly by the T4 DNA ligase ensuring thermal and physical stability of the resulting nano-structure.

As shown in figure 1-I, the designed tetrahedron is assembled from four oligonucleotides (labelled 1^t to 4^t), each containing three three-thymidine stretches spacing three edge-forming sequences (1^t A, 1^t B, 1^t C, 2^t A, etc.), which should each match specifically to one and only one complementary sequence on a partner oligonucleotide (for example the sequence of 1^t B should match the sequence of 2^t B and only that sequence) to allow efficient tetrahedron assembly. Hence, in total, the tetrahedron contains $4 \times 3 = 12$ base pairing sequences that should be pairwise complementary, and consequently six (since the sequence of one strand automatically determines the sequence of the complementary strand, the number of sequences is calculated by $(4 \times 3) / 2$) 18 nucleotides long sequences should be chosen and optimized using MIOS.

The cube (figure 1-II) is composed of six oligonucleotides (labelled 1^c to 6^c) each containing four three-thymidine stretches spacing four edge-forming sequences (1^c A, 1^c B, 1^c C, 1^c D, 2^c A, etc.), which should base pair specifically with complementary sequences on

partner oligonucleotides. Hence, 12 $((6 \times 4)/2)$ 18 nucleotides long sequences should be optimized by MIOS to allow assembly of this structure.

The octahedron (figure 1-III) consists of eight oligonucleotides (labelled 1° to 8°), each containing three edge-forming sequences (1° A, 1° B, 1° C, 2° A, etc.) spaced by three-thymidine stretches. Thus, for the octahedron assembly 12 $((8 \times 3)/2)$ 18 nucleotides long sequences should be chosen and optimized using MIOS. Note, that since both the cube and octahedron require 12 pairs of complementary sequences (a total of 24 pairwise complementary base pairing sequences) the same output sequences from MIOS were used for designing the oligonucleotides for assembly of both these structures.

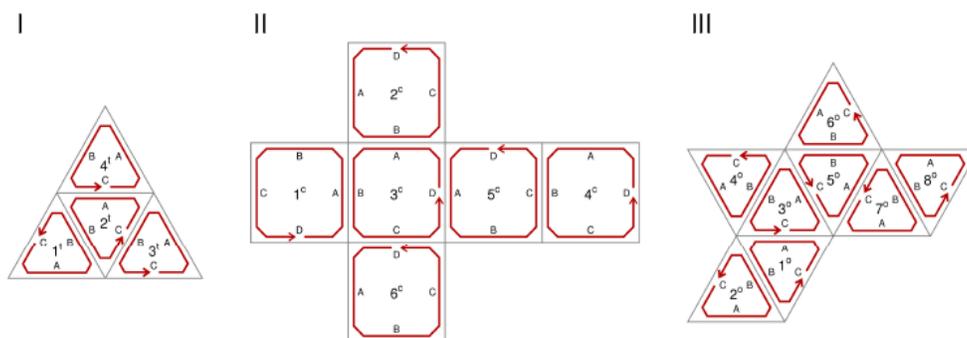


Figure 1. Folding models of a tetrahedron, a cube and an octahedron with truncated corners: All three structures consist of 18 base pair double-stranded edges interrupted by single-stranded stretches constituting the truncated corners. I, The tetrahedron consists of four oligonucleotides, 1^t to 4^t , each containing three edge-forming sequences (A, B and C). II, The cube consists of six oligonucleotides, 1^c to 6^c , each containing four edge-forming sequences (A, B, C and D). III, The octahedron consists of eight oligonucleotides, 1^o to 8^o , each containing three edge-forming sequences (A, B and C). The oligonucleotides are shown in red with the arrowheads designating their 3' end.

Optimization of DNA sequences: MIOS was written in “Python” and uses a greedy algorithm to create an optimized set of sequences for the design of oligonucleotides for one-step DNA nano-structure assembly by minimizing i) the sequence similarity between individual stretches of bases (as for example in the tetrahedron design, between the stretch of bases in 1^t A (see figure 1) and any other stretch of bases (e.g. 1^t B, 1^t C, 2^t A, etc. in figure 1)) as well as ii) the presence of complementary sequences within the same stretch of bases (for example within 1^t A, figure 1). As input to MIOS the desired number, length, and GC content of sequences are defined. Based on this input information, the software creates a random starting set of sequences. Thereafter, it identifies the worst sequence (with regards to points i) and ii) mentioned above) and replaces it with a better sequence if possible. MIOS repeats this process iteratively according to a user-defined threshold. In other words, the greedy algorithm replaces the worst sequence with the first random sequence that is found to have a lower maximum similarity to the rest of the sequence set than its predecessor. The software terminates when it has not decreased the global maximum similarity or global average similarity of the entire set of sequences in a specified number of iterations. MIOS outputs the progress of the run as well as the initial and optimized sets of sequences.

For each of the three structures experimentally tested below, three independent MIOS runs were carried out and the best set of sequences was chosen. The MIOS input

configurations of the edge forming sequences for the three structures were: the tetrahedron, 6 sequences, length = 18 base pairs, GC content = 0.5; the cube and the octahedron, 12 sequences, length = 18 base pairs, GC content = 0.5. For all runs, a threshold of 100,000 iterations without decreased similarity (i.e. improved annealing specificity) was chosen. For comparison to random guessing, one million random sequence sets were sampled for all three structures.

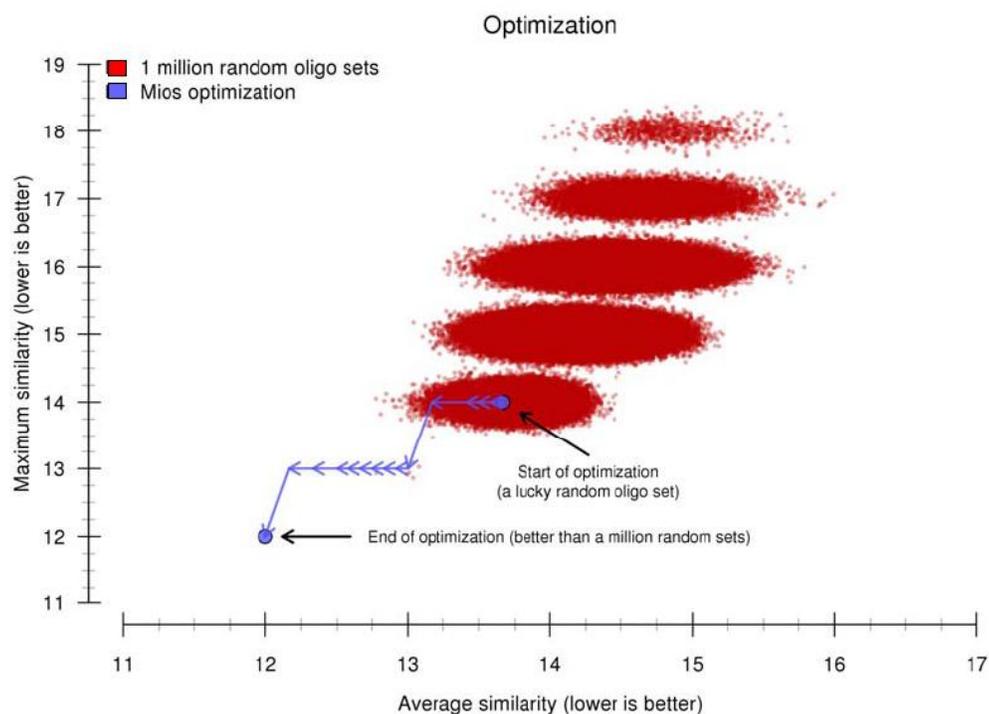


Figure 2. Graphical representation of MIOS optimization from a randomly guessed starting point. The MIOS run for the “cube/octahedron” sequences is shown in blue, with a starting point at 13.7 (average similarity, abscissas) and 14.0 (maximum similarity, ordinate). Within a few hundred iterations the algorithm decreases the maximum and average similarities to ~ 13.0 . After further iterations the algorithm finally reaches the endpoint at ~ 12.0 for the maximum and average similarities. A million random guesses (red dots) all have much higher average and maximum similarities. The random guesses have been spread out in order to visualize the density.

As shown in figure 2 for optimization of sequences for the cube/octahedron nanostructures, the MIOS output sequence sets were all highly optimized when compared to the distribution of one million random guesses and the software quickly achieved both a maximum- and an average similarity lower than the best of the million guesses. The optimization of sequence sets for the three structures resulted in the output sequences shown in table 1 (note the sequence sets used for the cube and the octahedron construction were identical). Sequences complementary to the output sequences were generated manually and the composition of each oligonucleotide for the assembly of the three structures with regards to their edge-forming (optimized sequences) and corner-forming (three thymidines) sequences determined using the folding models (Figure 1). The resulting oligonucleotide sequences used for the assemblies are shown in table 2.

Assembly and analysis of tetrahedral, cubic and octahedral DNA nano-structures built from oligonucleotides based on MIOS optimized sequences: The specific one-step assembly into tetrahedral, cubic, or octahedral nano-structures from oligonucleotides based on MIOS optimized sequences was tested by analyzing the products obtained when annealing and ligating increasing numbers of equimolar amounts of the designed oligonucleotides (table 2) in 5% native polyacrylamide gels essentially as described previously (Andersen *et al.*, 2008b, Oliveira *et al.*, 2010). As evident from figure 3 successive addition of oligonucleotides, designed to form each of the structures, to the assembly reaction resulted in specific products with decreasing mobility in the native gel. This electrophoretic pattern supports the formation of a tetrahedron, cube or octahedron using oligonucleotides based on MIOS optimized sequences.

Table 1. MIOS output sequences and the inverse complements used for the assembly of the tetrahedron, cube and octahedron. The name (seq name) of each MIOS output sequence (third column) was chosen to match the nomenclature of figure 1. Hence, sequence 1¹ B_{op} is the MIOS optimized version of sequence 1¹ B of figure 1 and so on. The sequences shown in the fifth column are the complementary partners of the sequences shown in the third column

	Output sequences (shown 5'-3')		Complementary sequences (shown 5'-3')	
	seq name	base sequence	seq name	base sequence
Tetrahedron	1 ¹ B _{op}	GAGCGGAGATCTATCTTG	2 ¹ B _{op}	CAAGATAGATCTCCGCTC
	1 ¹ C _{op}	CAACCACGATCTGAAGAC	4 ¹ B _{op}	GTCTTCAGATCGTGGTTG
	2 ¹ C _{op}	CCTTGTTGATCTGTGCAG	3 ¹ B _{op}	CTGCACAGATCAACAAGG
	3 ¹ A _{op}	CTTCTCCGATCTTCTTC	4 ¹ A _{op}	GGAAGAAGATCGGAGAAG
	3 ¹ C _{op}	GACTTAAGATCAGGCCAC	1 ¹ A _{op}	GTGGCCTGATCTTAAGTC
	4 ¹ C _{op}	CAGAGTGGATCAAGGATG	2 ¹ A _{op}	CATCCTTGATCCACTCTG
Cube	1 ^c A _{op}	GATATGTGATCCGCTGTC	3 ^c B _{op}	GACAGCGGATCACATATC
	1 ^c D _{op}	GCCTAATGATCACGAAGG	6 ^c A _{op}	CCTTCGTGATCATTAGGC
	2 ^c A _{op}	GTAGAACGATCGGAGAAC	1 ^c B _{op}	GTTCTCCGATCGTTCTAC
	2 ^c D _{op}	GGAGTACGATCCAACATG	4 ^c A _{op}	CATGTTGGATCGTACTCC
	3 ^c A _{op}	GCAAGGAGATCCTGTTAG	2 ^c B _{op}	CTAACAGGATCTCCTTGC
	3 ^c D _{op}	CTCGCAAGATCCATAACC	5 ^c A _{op}	GGTTATGGATCTTGCGAG
	4 ^c C _{op}	GATTACAGATCGGTGGTG	6 ^c B _{op}	CACCACCGATCTGTAATC
	4 ^c D _{op}	GAGGATTGATCAAGACCG	1 ^c C _{op}	CGGTCTTGATCAATCCTC
	5 ^c C _{op}	GGTGGTCGATCTATTGTC	4 ^c B _{op}	GACAATAGATCGACCACC
	5 ^c D _{op}	CTATGAGGATCTGAGTGG	2 ^c C _{op}	CCACTCAGATCCTCATAG
	6 ^c C _{op}	CGAAGAAGATCATGGTCC	5 ^c B _{op}	GGACCATGATCTTCTTCG
	6 ^c D _{op}	CTCTCTTGATCCTGTTCCG	3 ^c C _{op}	CGAACAGGATCAAGAGAG
Octahedron	1 ^o B _{op}	GATATGTGATCCGCTGTC	2 ^o B _{op}	GACAGCGGATCACATATC
	1 ^o C _{op}	GGTGGTCGATCTATTGTC	7 ^o A _{op}	GACAATAGATCGACCACC
	2 ^o C _{op}	CTATGAGGATCTGAGTGG	4 ^o A _{op}	CCACTCAGATCCTCATAG
	3 ^o B _{op}	GTAGAACGATCGGAGAAC	4 ^o B _{op}	GTTCTCCGATCGTTCTAC
	3 ^o C _{op}	CGAAGAAGATCATGGTCC	1 ^o A _{op}	GGACCATGATCTTCTTCG
	4 ^o C _{op}	CTCGCAAGATCCATAACC	6 ^o A _{op}	GGTTATGGATCTTGCGAG
	5 ^o B _{op}	GCAAGGAGATCCTGTTAG	6 ^o B _{op}	CTAACAGGATCTCCTTGC
	5 ^o C _{op}	GCCTAATGATCACGAAGG	3 ^o A _{op}	CCTTCGTGATCATTAGGC
	6 ^o C _{op}	CTCTCTTGATCCTGTTCCG	8 ^o A _{op}	CGAACAGGATCAAGAGAG
	7 ^o B _{op}	GATTACAGATCGGTGGTG	8 ^o B _{op}	CACCACCGATCTGTAATC
	7 ^o C _{op}	GAGGATTGATCAAGACCG	5 ^o A _{op}	CGGTCTTGATCAATCCTC
	8 ^o C _{op}	GGAGTACGATCCAACATG	2 ^o A _{op}	CATGTTGGATCGTACTCC

Table 2. Oligonucleotide sequences for assembly of the tetrahedron, cube, and octahedron. Based on the folding models shown in figure 1, the MIOS output sequences of table 1 were combined to the oligonucleotide sequences shown in the third column.

The corner-forming three-thymidines stretches are underlined>. The names of the oligonucleotides (OL name) were chosen to match the nomenclature of figure 1. For example oligonucleotide 1^t for the tetrahedron structure is named 1^t_{op} to denote that this particular oligonucleotide has sequences optimized by MIOS

	OL name	Oligonucleotides (shown 5'-3') OL sequence
Tetrahedron	1 ^t _{op}	TCTGAAGACTTTGTGGCCTGATCTTAAGTCTTTGAGCGGAGATCTATCTTGTTC AACCACGA
	2 ^t _{op}	TCTGTGCAGTTTCATCCTTGATCCACTCTGTTTCAAGATAGATCTCCGCTCTTCCCTTGTGA
	3 ^t _{op}	TCAGGCCACTTTCTTCTCCGATCTTCTTCTTCTGACAGATCAACAAGGTTTGACTTAAGA
	4 ^t _{op}	TCAAGGATGTTTGGAGAAGATCGGAGAAGTTTGTCTTCAGATCGTGGTTGTTTCAGAGTGGA
Cube	1 ^c _{op}	TCACGAAGGTTTGATATGTGATCCGCTGTCTTTGTTCTCCGATCGTTCTACTTTCCGGTCTTGATCA ATCCTCTTTGCCAATGA
	2 ^c _{op}	TCCAACATGTTTGTAGAACGATCGGAGAAGTTTCTAACAGGATCTCCTTGC TTTCCACTCAGATC CTCATAGTTTGGAGTACGA
	3 ^c _{op}	TCCATAACCTTTGCAAGGAGATCCTGTTAGTTTGACAGCGGATCACATATCTTTGGAACAGGATC AAGAGAGTTTCTCGAAGA
	4 ^c _{op}	TCAAGACCGTTTCATGTTGGATCGTACTCCTTTTGACAATAGATCGACCACCTTTGATTACAGATC GGTGGTGTITGAGGATTGA
	5 ^c _{op}	TCTGAGTGGTTTGGTTATGGATCTTGCAGTTTGGACCATGATCTTCTTCGTTTGGTGGTGCATC TATTGCTTTCTATGAGGA
	6 ^c _{op}	TCCTGTTCTGTTTCTTCTGATCATTAGGC TTTACCACCGATCTGTAATCTTTGGAAGAAGATCA TGGTCC TTTCTCTTTGA
Octahedron	1 ^o _{op}	TCTATTGCTTTGGACCATGATCTTCTTCGTTTGATATGTGATCCGCTGTCTTTGGTGGTCCGA
	2 ^o _{op}	TCTGAGTGGTTTCATGTTGGATCGTACTCCTTTGACAGCGGATCACATATCTTCTATGAGGA
	3 ^o _{op}	TCATGGTCC TTTCTTCTCGTATCATTAGGCTTTGTAGAACGATCGGAGAAGTTTTCGAAGAAGA
	4 ^o _{op}	TCCATAACCTTTCCACTCAGATCCTCATAGTTTGTCTCCGATCGTTCTACTTTCTCGAAGA
	5 ^o _{op}	TCACGAAGGTTTCGGTCTTGATCAATCCTTTTGC AAGGAGATCCTGTTAGTTTGCCTAATGA
	6 ^o _{op}	TCCTGTTCTGTTTGGTTATGGATCTTGCAGTTTCTAACAGGATCCTTGC TTTCTCTCTTGA
	7 ^o _{op}	TCAAGACCGTTTGACAATAGATCGACCACCTTTGATTACAGATCGGTGGTGTITGAGGATTGA
	8 ^o _{op}	TCCAACATGTTTTCGAACAGGATCAAGAGAGTTTACCACCGATCTGTAATCTTTGGAGTACGA

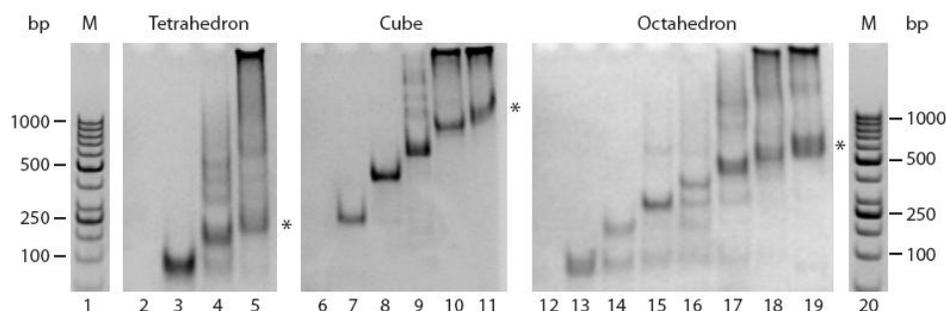


Figure 3. Gel-electrophoretic analyses of the partly and fully assembled DNA nano-structures: Lanes 1 and 20 show the mobility of a molecular marker consisting of 1000, 900, 800, 700, 600, 500, 400, 300, 250, 200, and 100 base pair double-stranded DNA fragments. Lanes 2-5 show the results of subjecting assembly reactions containing oligonucleotides 1^t_{op}, 1^t_{op} and 2^t_{op}, 1^t_{op}-3^t_{op} or 1^t_{op}-4^t_{op} to analysis in a native polyacrylamide gel. Lanes 6-11 show the results of subjecting assembly reactions containing 1^c_{op}, 1^c_{op} and 2^c_{op}, 1^c_{op}-3^c_{op}, 1^c_{op}-4^c_{op}, 1^c_{op}-5^c_{op} or 1^c_{op}-6^c_{op} to analysis in a native polyacrylamide gel. Lanes 12-19 show the results of subjecting assembly reactions containing 1^o_{op}, 1^o_{op} and 2^o_{op}, 1^o_{op}-3^o_{op}, 1^o_{op}-4^o_{op}, 1^o_{op}-5^o_{op}, 1^o_{op}-6^o_{op}, 1^o_{op}-7^o_{op} or 1^o_{op}-8^o_{op} to analysis in a native polyacrylamide gel. The gel-electrophoretic products corresponding to the fully assembled structure are marked by asterisks. The sizes of selected marker bands (in base pairs (bp)) are shown to the left and right of the gel-picture.

As estimated from the gel-electrophoretic mobilities, the products in figure 3, lanes 5, 11, and 19 (marked by asterisks) represent fully assembled tetrahedron, cube and octahedron, respectively. The assembly yields of the three structures were estimated to 25%, 33%, and 33%, respectively, by quantification of the fully assembled products, relative to the unspecific high and/or low mobility products in the gels.

ATOMISTIC MODELLING FOR VALIDATION OF DNA NANO-STRUCTURES

Atomistic modelling such as energy minimization and molecular dynamics (MD) simulation may be of great value for validating DNA nano-structures with respect to assembly probability and structural properties before experiments. The structure of any DNA configuration can be optimized by energy minimization and/or MD simulations mainly because of the development of reliable molecular force fields, as the Amber 99 (Cheatham *et al.*, 1999) implemented with the parmbsc0 corrections (Perez *et al.*, 2007) that permits the evaluation of the energetic contribution of any atoms belonging to a DNA molecule.

Energy minimization as a tool to predict assembly of octahedral DNA nano-cages: The approach of energy minimization optimizes the stability of the analyzed structure and identifies, at the atomistic level, the interactions responsible for possible instabilities. The power of the method has recently been demonstrated by determining the relative stability of octahedral DNA nano-cages, engineered using the structural framework outlined in figure 1, and differing one from another only by the length of the single-stranded thymidine stretches, which bridge the DNA double-helical edges of identical length (Oliveira *et al.*, 2010). In this example the atomistic model was built by placing the center of 12 double-stranded DNA helices at an identical distance from the center of a hypothetical octahedron. The relative orientations of the helices were optimized by minimizing the distance of the terminal atoms of the helices to the value corresponding to the length of the single-stranded thymidine stretches that bind to them and that vary depending on the number of thymidines (see figure 4). Once the optimal DNA helix orientations were reached, the thymidine linkers were manually added using the program SYBYL (TRIPOS, <http://www.tripos.com/>), removing the clashes through the SYBYL anneal module. Subsequently, each cage structure was optimized through several steps of energy minimization (using the generalized Born solvent model (Hawkins *et al.*, 1996) implemented in the program AMBER 9.0 (Case *et al.*, 2005) using the AMBER03 Force Field (Ponder and Case, 2003). The structure and energy of the final cages were then compared. Analysis, through the program *deform_energy* (<http://3dna.rutgers.edu>) (Lankas *et al.*, 2003), of the hydrogen bond deformation energy between each base pair in all the 12 double-stranded arms forming the different octahedral cages, permitted measurements of the deformations induced by the minimization algorithm. An optimal value was reached in each considered cage for the base pairs located in the middle part of the double-stranded helices. The results of this analysis showed that in the cages with three-, five- or seven-thymidine stretches the extremities of the double-stranded arms are slightly deformed showing average deformation energies of about 2.0 kcal/mol, whereas in the cage with two-thymidine stretches the extremities are much more deformed, reaching an average deformation energy of about 7.0 kcal/mol, with some peaks higher than 15.0 kcal/mol. This result indicated that the weak

point of this nano-structure is localized at the extremities of the double-stranded helices where a high local distortion does not permit the assembly of the two-thymidine stretch nano-cage. This is actually what was experimentally observed (Oliveira *et al.*, 2010). Such a high deformation is not present in the nano-cages having longer thymidine stretches providing an energetic explanation for the reasonably high experimental yields (~30%) observed for these cages. Hence, the length of the thymidine stretch appears to be the crucial parameter for efficient assembly of the DNA octahedral cages, with three nucleotides being the minimum length required for an optimal pairing of the base pairs in the double-stranded helices.

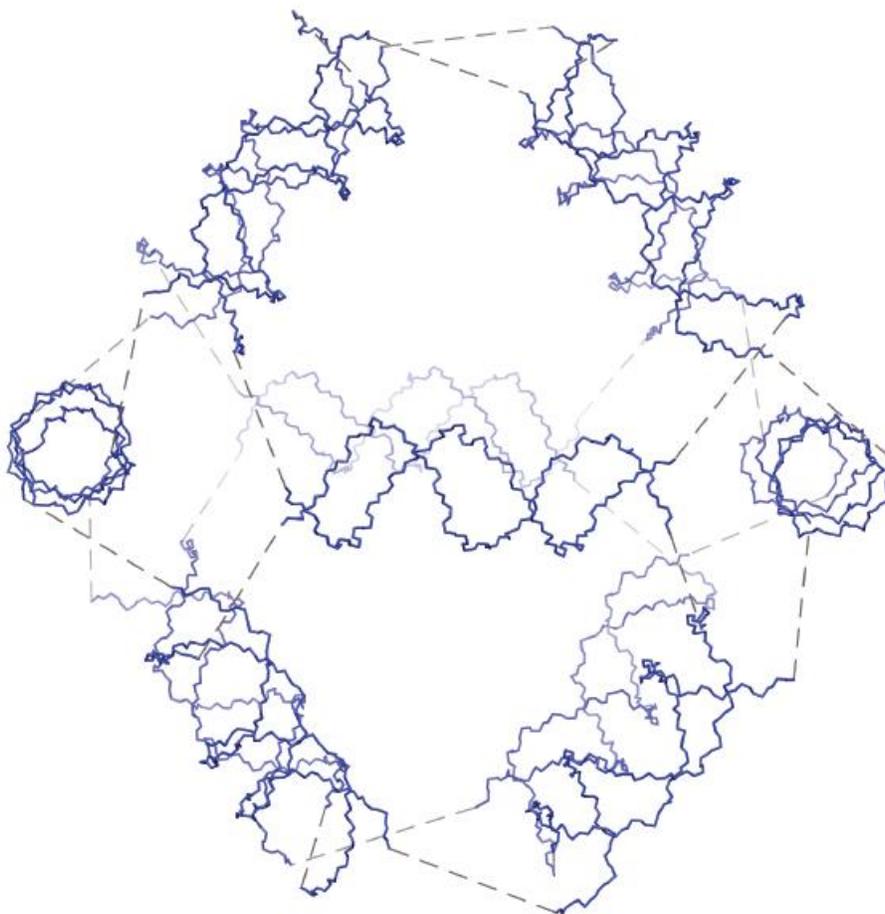


Figure 4. 3D arrangement of the 12 double-stranded helices in an octahedral geometry. The dashed lines represent the single-stranded thymidine stretches, of which the length modulates the inter-helices distance.

The value of MD simulations in DNA nano-science: The link between structure and dynamics of DNA nano-structures can be obtained by extensive classical MD simulations enabling the exploration of the conformational energy landscape accessible to the investigated structures. The complexity of biological model systems studied by MD simulations has increased dramatically since the first reported simulation of a small protein (McCammon *et al.*, 1977), allowing the exploration of many of the fast phenomena that occur within proteins

and DNA at the atomistic level. As a matter of fact, MD simulation of proteins in a water box has become a relatively routine approach and reliable results have been obtained for the past 20 years. In the case of DNA, reliable MD simulations have been carried out for a decade (Bonvin *et al.* 1998, Castrignanò *et al.* 2000) permitting scientists to obtain detailed description of protein-DNA recognition at the atomic level. Although the technique has only recently been applied to the descriptions of synthetic DNA nano-structures, it is predictable that the wealth of information provided by MD simulations on such structures most likely will make it a commonly used tool in DNA nano-science.

MD simulation has been used to evaluate the structure and stability of several paranemic crossover DNA structures, where analysis of the vibrational mode has permitted the identification of the most stable structure (Maiti *et al.*, 2006). A more recent application concerns the study of the previously described octahedral DNA nano-cage with seven-thymidine single-stranded stretches. The simulation of this nano-structure was done on an atomistic model immersed in a truncated octahedral box filled with TIP3P water molecules, imposing a minimal distance between the solute and the box walls of 10.0 Å. The system was neutralized through the AMBER leap module, adding 600 Na⁺ ions in electrostatically favorable positions. The final system consisted of 392,955 atoms. After relaxation of the entire system through energy minimization and MD simulation of the solvent and the ions, the thymidine stretches were equilibrated through energy minimization and MD simulation, restraining the atoms of the DNA double-stranded helices with a 500 kcal force constant. The whole solute was minimized by imposing a 20 kcal force constant and finally simulated for 40 ps with 2.0 fs time steps, without any restraint, at a constant temperature of 300 K, using the Berendsen's method, and at a constant pressure of 1 bar. After this procedure the system was simulated for 12 ns. Analysis of the simulation indicated that the DNA double-stranded helices, composing the edges of the nano-cage structure, are very stable and that the single-stranded thymidine stretches give a significant contribution to the organization of the scaffold geometry. Taken together, the results of the MD simulation suggest that the global scaffold of the nano-cage undergoes a contraction and that the length of the thymidine stretches is the crucial aspect in the modulation of the nano-cage stability (Falconi *et al.*, 2009). Hence, from the MD simulations it was concluded that modulation of the length and nucleotide composition of the single-stranded stretches connecting the double-stranded helices determines the rigidity of the DNA nano-cages, giving the opportunity to engineer cages with varied flexibility.

ACKNOWLEDGMENT

We are grateful to BSc. Esben Eickhardt for scientific assistance. The study was supported by the Foundation til Lægevidenskabens fremme, the Augustinus Foundation, the Civilingeniør Frode V. Nyegaard og hustrus Foundation, the Fabrikant Einar Willumsens Foundation and the Købmand Knud Øster-Jørgensen og Maria Øster-Jørgensens Foundation.

REFERENCES

- Aldaye, F. A., and Sleiman, H. F. (2007) Modular access to structurally switchable 3D discrete DNA assemblies. *J. Am. Chem. Soc.*, 129, 13376-13377.
- Andersen, E. S., Dong, M., Nielsen, M. M., Jahn, K., Lind-Thomsen, A., Mamdouh, W., Gothelf, K. V., Besenbacher, F., and Kjems, J. (2008a) DNA origami design of dolphin-shaped structures with flexible tails. *ACS Nano.*, 2, 1213-1218.
- Andersen, E. S., Dong, M., Nielsen, M. M., Jahn, K., Subramani, R., Mamdouh, W., Golas, M. M., Sander, B., Stark, H., Oliveira, C. L., Pedersen, J. S., Birkedal, V., Besenbacher, F., Gothelf, K. V., and Kjems, J. (2009) Self-assembly of a nanoscale DNA box with a controllable lid. *Nature*, 459, 73-76.
- Andersen, F. F., Knudsen, B., Oliveira, C. L., Fröhlich, R. F., Krüger, D., Bungert, J., Agbandje-McKenna, M., McKenna, R., Juul, S., Veigaard, C., Koch, J., Rubinstein, J. L., Guldbrandtsen, B., Hede M. S., Karlsson, G., Andersen, A. H., Pedersen, J. S., and Knudsen, B. R. (2008b) Assembly and structural analysis of a covalently closed nanoscale DNA cage. *Nucleic Acids Res.*, 36, 1113-1119.
- Birac, J. J., Sherman, W. B., Kopatsch, J., Constantinou, P. E., and Seeman, N. C. (2006) Architecture with GIDEON, a program for design in structural DNA nanotechnology. *J. Mol. Graph. Model.*, 25, 470-80.
- Bhatia, D., Mehtab, S., Krishnan, R., Indi, S. S., Basu, A., and Krishnan, Y. (2009) Icosahedral DNA nanocapsules by modular assembly. *Angew Chem. Int. Ed. Engl.* 48, 4134-4137.
- Bonvin, A. M., Sunnerhagen, M., Otting G., and van Gunsteren, W. F. (1998) Water molecules in DNA recognition. A molecular dynamics view of the structure and hydration of the trp operator. *J. Mol. Biol.*, 282, 859-873.
- Castrignanò, T., Chillemi, G., and Desideri, A. (2000) Structure and Hydration of BamHI DNA Recognition Site: A Molecular Dynamics Investigation. *Biophys. J.*, 79, 1263-1272.
- Case, D. A., Cheatham III, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., Onufriev Jr., A., Simmerling, C., Wang, B., and Woods, R. (2005) The AMBER Biomolecular Simulation Programs. *J. Comput. Chem.*, 26, 1668-1688.
- Cheatham, T. E., Cieplak, P., and Kollman, P. A. (1999) A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat. *J. Biomol. Struct. Dyn.*, 16, 845-862.
- Chen, J. H., and Seeman, N.C. (1991) Synthesis from DNA of a molecule with the connectivity of a cube. *Nature.*, 350, 631-633.
- Deaton, R., Kim, J. W., and Chen, J. (2003) Design and test of non-crosshybridizing oligonucleotide building blocks for DNA computers and nanostructures. *Applied Physics Letters*, 82, 1305-1307.
- Douglas, S. M., Dietz, H., Liedl, T., Högberg, B., Graf, F., and Shih, W. M. (2009a) Self-assembly of DNA into nanoscale three-dimensional shapes. *Nature*, 459, 414-418.
- Douglas, S. M., Marblestone, A. H., Teerapittayanon, S., Vazquez, A., Church, G. M., and Shih, W. M. (2009b) Rapid prototyping of 3D DNA-origami shapes with caDNAno. *Nucleic Acids Res.* 37, 5001-5006.

- Falconi, M., Oteri, F., Chillemi, G., Andersen, F. F., Tordrup, D., Oliveira, C. L., Pedersen, J. S., Knudsen, B. R., and Desideri, A. (2009) Deciphering the Structural Properties That Confer Stability to a DNA Nanocage. *ACS Nano.*, 3, 1813–1822.
- Goodman, R. P. (2005) NANEV: a program employing evolutionary methods for the design of nucleic acid nanostructures. *Biotechniques.* 38, 548-550.
- Goodman, R. P., Schaap, I. A., Tardin, C. F., Erben, C. M., Berry, R. M., Schmidt, C. F., and Turberfield, A. J. (2005) Rapid chiral assembly of rigid DNA building blocks for molecular nanofabrication. *Science.* 310, 1661-1665.
- Hawkins, G. D., Cramer, C. J., and Truhlar, D. G. (1996) Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. *J. Phys. Chem.*, 100, 19824-19839.
- He, Y., Ye, T., Su, M., Zhang, C., Ribbe, A. E., Jiang, W., and Mao, C. (2008) Hierarchical self-assembly of DNA into symmetric supramolecular polyhedra. *Nature*, 452, 198-201.
- Kato, T., Goodman, R. P., Erben, C. M., Turberfield, A. J., and Namba, K. (2009) High-resolution structural analysis of a DNA nanostructure by cryoEM. *Nano Lett.*, 9, 2747-2750.
- Kumara, M. T., Nykypanchuk, D., and Sherman, W. B. (2008) Assembly pathway analysis of DNA nanostructures and the construction of parallel motifs. *Nano Lett.* 8, 1971-1977.
- LaBean, T. H., Yan, H., Kopatsch, J., Liu, F., Winfree, E., Reif, J. H., and Seeman, N. C. (2000) Construction, Analysis, Ligation, and Self-Assembly of DNA Triple Crossover Complexes. *J. Am. Chem. Soc.* 122, 1848-1860.
- Lankas, F., Sponer, J., Langowski, J., and Cheatham, T. E. (2003) DNA basepair step deformability inferred from molecular dynamics simulations. *Biophysical journal*, 85, 2872-2883.
- Maiti, P. K., Pascal, T. A., Vaidehi, N., Heo, J., and Goddard, W. A. 3rd. (2006) Atomic-level simulations of seeman DNA nanostructures: the paranemic crossover in salt solution. *Biophys J.*, 90, 1463-1479.
- Mathieu, F., Liao, S., Kopatsch, J., Wang, T., Mao, C., and Seeman, N. C. (2005) Six-helix bundles designed from DNA. *Nano Lett.*, 5, 661-665.
- McCammon, J. A., Gelin, B. R., and Karplus, M. (1977) Dynamics of Folded Proteins. *Nature* 267, 585–590.
- Oliveira, C. L., Juul, S., Jørgensen, H. L., Knudsen, B., Tordrup, D., Oteri, F., Falconi, M., Koch, J., Desideri, A., Pedersen, J. S., Andersen, F. F., and Knudsen, B. R. (2010) Structure of nanoscale truncated octahedral DNA cages: variation of single-stranded linker regions and influence on assembly yields. *ACS Nano.*, 4, 1367-1376.
- Perez, A., Marchan, I., Svozil, D., Sponer, J., Cheatham, T. E., III, Lughton, C. A., and Orozco, M. (2007) Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys. J.*, 92, 3817–3829.
- Ponder, J. W., and Case, D. A. (2003) Force Fields for Protein Simulations. *Adv. Protein Chem.*, 66, 27–85.
- Rothemund, P. W. (2006) Folding DNA to create nanoscale shapes and patterns. *Nature.*, 440, 297-302.
- Shih, W. M., Quispe, J. D., and Joyce, G. F. (2004) A 1.7-kilobase single-stranded DNA that folds into a nanoscale octahedron. *Nature*, 427, 618-621.

-
- Voigt, N. V., Tørring, T., Rotaru, A., Jacobsen, M. F., Ravnsbaek, J. B., Subramani, R., Mamdouh, W., Kjems, J., Mokhir, A., Besenbacher, F., and Gothelf, K. V. (2010) Single-molecule chemical reactions on DNA origami. *Nat. Nanotechnol.*, 5, 200-203.
- Yan, H., LaBean, T. H., Feng, L., and Reif, J. H. (2003) Directed nucleation assembly of DNA tile complexes for barcode-patterned lattices. *Proc. Natl. Acad. Sci. USA.*, 100, 8103-8108.
- Zheng, J., Birktoft, J. J., Chen, Y., Wang, T., Sha, R., Constantinou, P. E., Ginell, S. L., Mao, C., and Seeman, N. C. (2009) From molecular to macroscopic via the rational design of a self-assembled 3D DNA crystal. *Nature*. 461, 74-77.