Structure of Nanoscale Truncated Octahedral DNA Cages: Variation of Single-Stranded Linker Regions and Influence on Assembly Yields

Cristiano Luis Pinto Oliveira,[†,#] Sissel Juul,[‡,#] Hanne Lærke Jørgensen,[‡] Bjarne Knudsen,[§] David Tordrup,[‡] Francesco Oteri,[∥] Mattia Falconi,[∥,⊥] Jørn Koch,[¶] Alessandro Desideri,[∥,⊥] Jan Skov Pedersen,[†] Felicie Faucon Andersen,[‡,*] and Birgitta Ruth Knudsen[‡,*]

[†]Department of Chemistry and Interdisciplinary Nanoscience Center (iNANO), University of Aarhus, 8000 Aarhus C, Denmark, [‡]Department of Molecular Biology and Interdisciplinary Nanoscience Center (iNANO), Aarhus University, C.F. Møllers Allé, Bldg. 130, 8000 Aarhus C, Denmark, [§]CLC bio, Finlandsgade 10-12, 8200 Aarhus N, Denmark, [∥]Department of Biology and Center of Biostatistics and Bioinformatics, University of Rome "Tor Vergata", Via della Ricerca Scientifica 1, 00133 Rome, Italy, [⊥]NAST Nanoscience & Nanotechnology & Innovative Instrumentation, University of Rome "Tor Vergata", Via della Ricerca Scientifica 1, 00133 Rome, Italy, and [¶]Department of Pathology, Aarhus Sygehus, Nørrebrogade 44, 8000 Aarhus C, Denmark. [#]These authors contributed equally to this work.

ABSTRACT The assembly, structure, and stability of DNA nanocages with the shape of truncated octahedra have been studied. The cages are composed of 12 double-stranded B-DNA helices interrupted by single-stranded linkers of thymidines of varying length that constitute the truncated corners of the structure. The structures assemble with a high efficiency in a one-step procedure, compared to previously published structures of similar complexity. The structures of the cages were determined by small-angle X-ray scattering. With increasing linker length, there is a systematic increase of the cage size and decrease of the twist angle of the double helices with respect to the symmetry planes of the cage structure. In the present study, we demonstrate the length of the single-stranded linker regions, which impose a certain degree of flexibility to the structure, to be the important determinant for efficient assembly. The linker length can be decreased to three thymidines without affecting assembly yield or the overall structural characteristics of the DNA cages. A linker length of two thymidines represents a sharp cutoff abolishing cage assembly. This is supported by energy minimization calculations suggesting substantial hydrogen bond deformation in a cage with linkers of two thymidines.

KEYWORDS: DNA nanostructure · single-stranded linker variation · assembly efficiency · flexibility · small-angle X-ray scattering

The unique self-recognition properties of DNA determined by the strict rules of Watson–Crick base pairing makes this material ideal for the creation of self-assembling predesigned nanostructures in a bottom-up approach.[1] The construction of such structures is one of the main focuses of the thriving area of DNA nanotechnology, where several assembly strategies have been employed to build increasingly complex three-dimensional (3D) DNA nanostructures.[2−13] During recent years, 3D structures including a truncated octahedron,[4] a tetrahedron, dodecahedron, and a DNA buckyball[8] have been assembled using double or multiple crossover DNA motifs. These structures are characterized by an impressively high assembly efficiency (typically around 90%) and a fairly high mechanical rigidity, which is ascribed to the complex crossover motifs by which they are built. For some purposes, however, the DNA crossover motifs may impose limitations to these structures in terms of their lack of covalent closure (which may leave them sensitive toward enzymatic or thermal degradation) as well as their expected limited reactivity with naturally occurring enzymes and proteins. Nature provides an extensive toolbox of DNA binding, cutting, ligating, or recombining enzymes, which may all prove valuable for synthesis, manipulation, or functionalization of DNA nanostructures. Therefore, although sometimes hampered by lower rigidity and/or assembly efficiency than structures built from multiple crossover motifs,[1] 3D structures mainly or entirely composed of double-stranded B-DNA[2,3,5−7,12,13] (the most common natural conformation of DNA) may offer advantages in terms of the ease by which they can be manipulated by natural enzymes.

Seeman and co-workers were the first to use B-DNA to assemble 3D structures with the connectivities of a cube and a truncated octahedron, although with a quite low yield of approximately 1%.[2,3] More recently published nanostructures composed of B-DNA include different tetrahedra characterized by both high yield (95%) and

*Address correspondence to fa@mb.au.dk, brk@mb.au.dk.

Received for review October 28, 2009 and accepted January 29, 2010.

Published online February 10, 2010. 10.1021/nn901510v

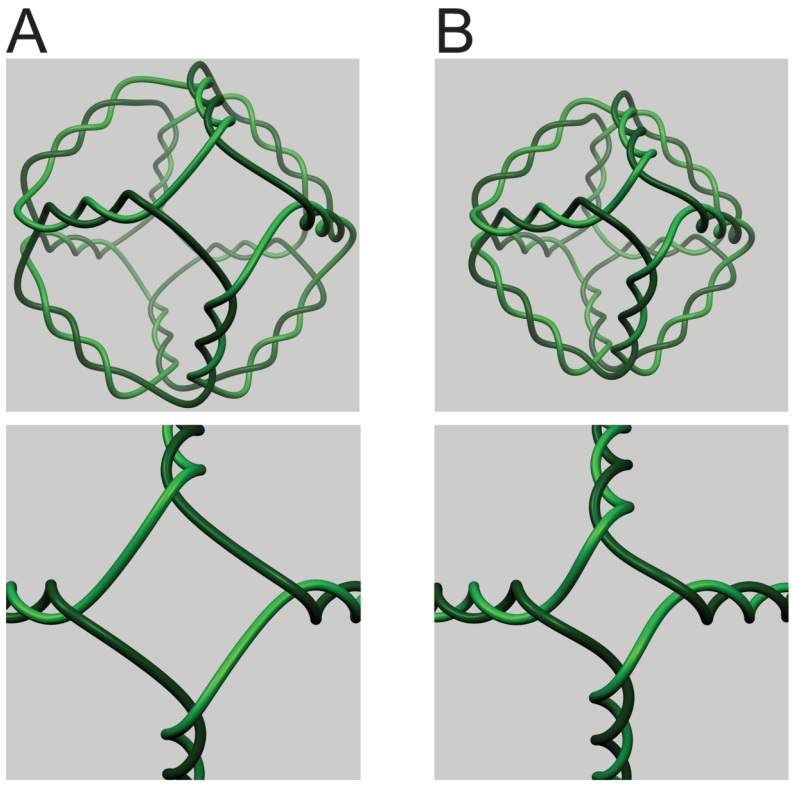
mechanical rigidity[6,12,14] and an icosahedron constructed by modular assembly and with the capability of encapsulating gold nanobeads.[13] An elegant approach of using simple units for the efficient assembly of numerous 3D structures including prisms of different geometries and DNA nanotubes with arms of B-DNA was presented by Aldaye and Sleiman.[5,15] The modular construction of DNA nanotubes[15] represents an important step toward the construction of DNA networks and lattices in two or three dimensions, which is one of the main goals of DNA nanotechnology.[1] Alternative routes to this goal may involve the reaction of relatively simple 3D units with natural DNA modifying enzymes, such as DNA topoisomerases or recombinases,[16−18] emphasizing the relevance of using natural DNA conformations in the buildup of 3D nanostructures.

We previously demonstrated the assembly of a covalently closed truncated octahedral cage composed of double-stranded B-DNA arms interrupted by single-stranded linkers of seven thymidines (denoted cage(7T) in the following) using a one-step assembly procedure.[7] When considering the complexity of the structure, the assembly yield of approximately 30% in a single step was surprisingly high. This high yield may be promoted by the single-stranded regions of the structure, which was suggested by recent molecular dynamics (MD) simulations to impose a certain degree of flexibility in the structure.[19] To investigate the important determinants of assembly yields within the basic structural framework of the originally presented cage(7T), we here address the effect of successively decreasing the single-stranded linker length (from seven to two thymidines) on cage assembly efficiency and structural properties. As a course of this study, we found that the single-stranded regions of the cage can be decreased from seven to three thymidines without affecting the yield or overall structure, as determined by small-angle X-ray scattering (SAXS). Further decreasing the linker length completely abolishes cage assembly most probably due to deformation of hydrogen bonds at the ends of the double-stranded arms as suggested by energy minimization calculations.

RESULTS AND DISCUSSION

Design of the DNA Cages: We previously reported the design and assembly of a DNA cage(7T) (schematically represented in Figure 1A) from eight 75-mer oligonucleotides, which upon annealing formed a truncated octahedron composed of 12 double-stranded regions of each 18 base pairs interrupted by 24 single-stranded linker regions of seven thymidines.[7] This structure assembled in a one-step procedure with a fairly high efficiency (around 30%) that may be promoted by a relatively high flexibility introduced by the single-stranded linkers as suggested by MD simulation.[19] To determine the effect of linker length on assembly efficiency, we designed five new sets of eight oligonucleotides identical to the ones assembling into cage(7T), except that the linker regions were successively shortened to form smaller and presumably less flexible cage structures (Figure 1B). The oligonucleotides designed to assemble into cage structures with six, five, four, three, or two thymidine linker regions, denoted cage(6T), cage(5T), cage(4T), cage(3T), or cage(2T) in the following, are shown in Figure 1C.

Assembly, Gel Electrophoretic Analysis, and Yield Estimation of DNA Cage(7T) to Cage(2T): Covalently closed truncated octahedral DNA cages with linker regions varying from seven to two thymidines (cage(7T) to cage(2T)) were assembled essentially as described previously.[7] The specific assembly of each DNA cage was tested by analyzing the products obtained when annealing and ligating increasing numbers of equimolar amounts of 5′-phosphorylated oligonucleotides added one-by-one in successive order from $OL1_n$−$OL8_n$ (see Figure 1C) to the assembly mixture in native polyacrylamide gels. The results of such experiments using oligonucleotides with seven, five, three, or two thymidine linkers are shown in Figure 2. As evident from Figure 2A−C, adding increasing numbers of oligonucleotides with stretches of either seven, five, or three thymidines to the assembly reaction resulted in products running mainly as single bands with decreasing mobility in the native gels. This gel electrophoretic pattern is consistent with already published results demonstrating specific assembly of cage(7T)[7] and supports the specific formation of the DNA cage(5T) and cage(3T). Moreover, the specific products observed in lanes 8 of Figure 2A−C were resistant toward boiling and denaturing gel electrophoresis (data not shown), suggesting that these products represent covalently closed cage structures. Similar results were obtained when testing the assembly of oligonucleotides $OL1_4$−$OL8_4$ or $OL1_6$−$OL8_6$ into cage(4T) and cage(6T), respectively (data not shown). In contrast, specific products with successively decreasing mobility were observed only when mixing $OL1_2$ with $OL2_2$, $OL1_2$−$OL3_2$, and $OL1_2$−$OL4_2$ (Figure 2D, lanes 2−4), most probably corresponding to the formation of a double-stranded region between $OL1_2$ and $OL2_2$ in lane 2, between $OL1_2$ and $OL2_2$ as well as between $OL1_2$ and $OL3_2$ in lane 3, and between $OL1_2$ and $OL2_2$, between $OL1_2$ and $OL3_2$, and between $OL3_2$ and $OL4_2$ in lane 4. No additional (higher mobility) products were observed upon addition of $OL5_2$ or $OL5_2$ and $OL6_2$ to the assembly mixtures (lanes 5 and 6), suggesting that annealing of $OL5_2$ and/or $OL6_2$ to the structure formed by $OL1_2$−$OL4_2$ is impaired when the single-stranded linker consists of two thymidines. The mixing of $OL1_2$−$OL7_2$ and $OL1_2$−$OL8_2$ resulted in unspecific high mobility products giving rise to a gel electrophoretic smear (lanes 7 and 8) probably caused by annealing of the oligonucleotides into large aggregates. This pattern was not changed upon changing the order by which



C

OL1$_n$:gccaccagg(t)$_n$cgatgtctaagctgaccg(t)$_n$ggaccgtgattccatgac(t)$_n$cttagagtt
OL2$_n$:tggctacag(t)$_n$cggtcagcttagacatcg(t)$_n$gaatcctatgctcggacg(t)$_n$ggctcacat
OL3$_n$:tcacggtcc(t)$_n$ctatccgatcgaggcatg(t)$_n$catactgagagcgttccg(t)$_n$gtcatggaa
OL4$_n$:cagatacgc(t)$_n$catgcctcgatcggatag(t)$_n$ctgtagccaatgtgagcc(t)$_n$gtcgcagtt
OL5$_n$:ctcagtatg(t)$_n$cggttacggtacaatgcc(t)$_n$cgcaagacgttagtgtcc(t)$_n$cggaacgct
OL6$_n$:ggtgtatcg(t)$_n$ggcattgtaccgtaaccg(t)$_n$gcgtatctgaactgcgac(t)$_n$ccaccgaat
OL7$_n$:cgtcttgcg(t)$_n$gtatgacgcagcacttgc(t)$_n$cctggtggcaactctaag(t)$_n$ggacactaa
OL8$_n$:ataggattc(t)$_n$gcaagtgctgcgtcatac(t)$_n$cgatacaccattcggtgg(t)$_n$cgtccgagc

Figure 1. Design of truncated octahedral DNA cages. (A) Schematic representation of cage(7T) (top panel) and an enlarged image of one of the single-stranded truncated vertices of the structure (bottom panel). (B) Schematic representation of cage(3T) (top panel) and an enlarged image of one of the single-stranded truncated vertices (bottom panel). (C) Sequences of the oligonucleotides used to assemble the cage(2T)−cage(7T). The specific sequences interrupted by thymidine linkers anneal pairwise to each other to form the double-stranded arms of the structures; t represents the linker regions, which remain single-stranded after assembly; n = 2, 3, 4, 5, 6, or 7 for oligonucleotides designed to form cage(2T), cage(3T), cage(4T), cage(5T), cage (6T), and cage(7T), respectively. The cartoons shown in A and B were generated using the open source tool for 3D creations, Blender software.

the oligonucleotides were added in the assembly mixture set (data not shown). Hence, $OL1_2$−$OL8_2$ are unable to assemble into a specific cage product, and three nucleotides, therefore, seem to represent the lower limit of the single-stranded linker region necessary for efficient cage assembly within the utilized structural framework. As estimated from quantification of the fully assembled products (lane 8, Figure 2A−C), relative to the unspecific high and/or low mobility products in gels obtained from three independent experiments (the gel pictures shown in Figure 2 being a typical representative of these), the approximate yields of cage(3T) to cage(7T) were all ~30% (data not shown). Hence, the difference of only one nucleotide in the linker, from three to two thymidines, appears to represent a sharp cutoff determining efficient or no assembly.

SAXS Analysis of DNA Cage(3T), Cage(4T), Cage(5T), Cage(6T), and Cage(7T): SAXS was used to investigate the structure of the cages with linker length varying from three to six thymidines, and the resulting structural models were compared to that of cage(7T) previously published.[7] The experimental data for the DNA cage(3T) to cage(7T) are shown in Figure 3A. This figure also shows the fitting using indirect Fourier transformation (IFT).[20] For all cages, it is possible to distinguish two regions in the experimental curve. The bumps at high q of the experimental curve demonstrate the presence of highly symmetric well-defined particles. The fits from the full curves (dashed lines in Figure 3A) furthermore indicate the presence of oligomers, varying from ~300 up to ~500 Å in size, as can be seen from the resulting $p(r)$ curves (Figure 3C). One tentative approach to eliminate

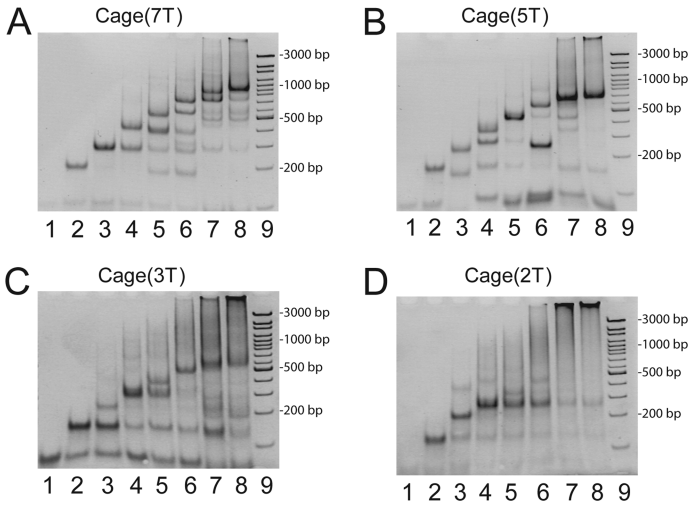


Figure 2. Gel electrophoretic analyses of the partly and fully assembled DNA cages. (A) Lanes 1−8 show the results of subjecting assembly reactions containing $OL1_n$, $OL1_n$, and $OL2_n$, $OL1_n$−$OL3_n$, $OL1_n$−$OL4_n$, $OL1_n$−$OL5_n$, $OL1_n$−$OL6_n$, $OL1_n$−$OL7_n$, or $OL1_n$−$OL8_n$ with $n = 7$ to analysis in a native polyacrylamide gel. (B−D) Lanes 1−8 are the same as in panel A, except that oligonucleotides with linker regions of five, three, or two thymidines were used. Lane 9 in all gel pictures is a molecular marker showing the mobility of double-stranded DNA of 3000, 2000, 1500, 1200, 1000, 900, 800, 700, 600, 500, 400, 300, 200, and 100 bp.

the effects of the aggregation into oligomers is to omit some points in the beginning of the experimental curve, as the high q part of the scattering data is dominated by scattering from the constituting cages. This procedure has intrinsic limitations, but if the oligomers are simple combinations of the monomer structure, without strong overlapping of the structures (oligomers built as side-by-side monomers), the suggested approach is almost valid.[7] The corresponding $p(r)$ functions for the monomeric cages are shown in Figure 3B.

TABLE 1. Summary of Parameters Obtained from the IFT Analysis[a]

	monomer region		full curve		
sample	R_g [Å]	D_{max} [Å]	R_g [Å]	D_{max} [Å]	$I(0)_{full}/I(0)_{cut}$
3T	57.2 ± 0.8	~148	86 ± 1	~296	1.6
4T	60.7 ± 0.4	~156	99 ± 4	~345	1.4
5T	62.2 ± 0.1	~163	128 ± 1	~374	2.8
6T	66.6 ± 0.7	~170	99 ± 4	~424	7.4
7T	67.5 ± 0.3	~165	129 ± 1	~380	3.3

[a]Summary of IFT analysis for the SAXS experimental data from cages with different thymidine linker lengths. Two regions were fitted for each curve. For the full curve, all points were considered, and for the "monomer region" curve, a few points were skipped at low q in order to eliminate the contribution from the oligomers. In each case, the radius of gyration (R_g) and maximum dimension (D_{max}) were obtained. Also, the forward scattering values for the two fits ($I(0)_{full}$ and $I(0)_{cut}$) were obtained, and the ratio between the value for the full curve fitting and the monomer region fitting was calculated.

The first bump at around 20 Å comes from the well-defined cross section of the double-stranded DNA, which builds up the particles. The values where the $p(r)$ functions go to zero at high r values directly give the maximum dimension D_{max} of the different DNA cages. All $p(r)$ functions have a maximum, which is shifted to higher values than $D_{max}/2$, which is characteristic of hollow particles. Figure 3D shows the radius of gyration R_g as a function of linker length. As could be expected, there is a consistent increase in the overall sizes of the cage radius of gyration as the linker length increases. A summary of the parameters obtained from the IFT analysis, radius of gyration R_g, maximum size D_{max}, and ratio of forward scattering for the full range fit $I(0)_{full}$ and fit for the monomeric part $I(0)_{cut}$, for the different cages is shown in Table 1. Since the forward scat-

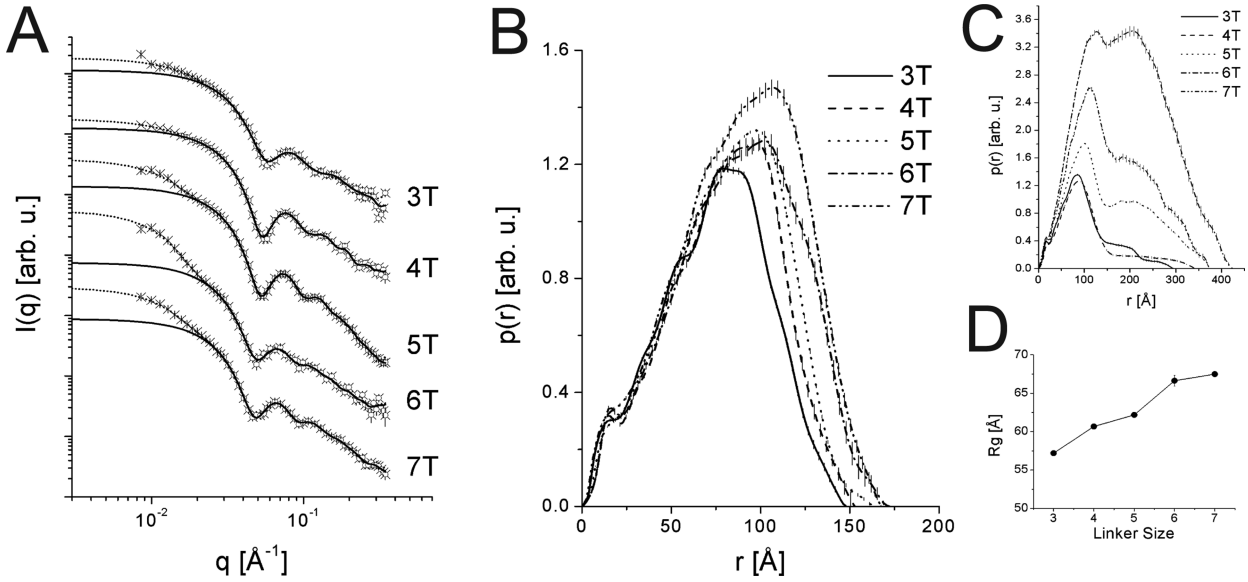


Figure 3. Indirect Fourier transformation analysis for the DNA cages. (A) SAXS experimental data for cage(3T), cage(4T), cage(5T), cage(6T), and cage(7T). Crosses: full experimental data for the studied samples. Circles: monomer contribution to the scattering data for each sample. Dashed lines: IFT fits for the full data set. Solid lines: IFT fits for the monomer region. The data sets were shifted for clarity. (B) Pair distance distribution functions $p(r)$ for the monomer region in each data set. (C) Pair distance distribution functions $p(r)$ for the full curve, providing information about the sizes of the oligomers in each sample. (D) Variation of the radius of gyration for each cage as a function of the linker length.

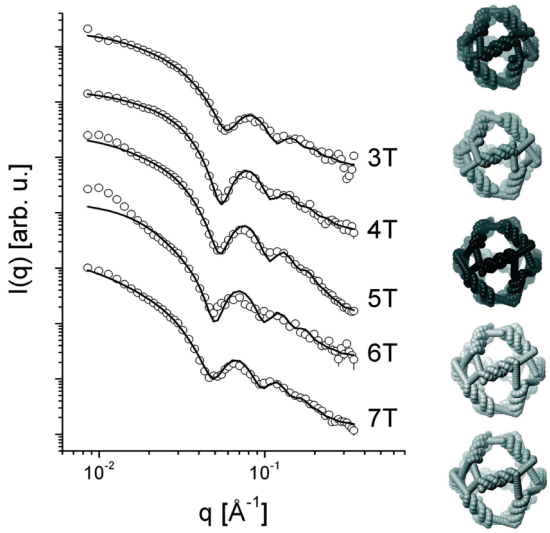


Figure 4. Fitting of the experimental data with the truncated octahedron model. Left: Fits of the experimental data for the samples with different thymidine linker lengths using the geometrical model. The data sets were shifted for clarity. Right: Resulting three-dimensional structures obtained from the modeling of the experimental data.

tering is proportional to the weight-average molecular weight of the sample, the ratio between the forward scattering for the full range fitting and the monomeric region allows estimation of an averaged number of monomers within the aggregated state. For samples with cage(3T) and cage(4T), we obtain values of 1.6 and 1.4, respectively, which shows that the samples are quite monodisperse with a quite small fraction of aggregates. For sample cage(5T), the data give a little higher oligomeric size with a ratio of 2.8, whereas for the sample cage(6T), the ratio 7.4 shows the presence of relatively large aggregates. The sample with cage(7T) has a ratio of 3.3, indicating smaller size oligomers, as previously reported.[7]

In order to obtain information about the population of sizes, we performed the modeling as described in the Methods section. The model consists of the cage structure as monomers and also arranged as dimers and trimers in linear arrangements. The structure of the cage as well as the populations of the monomers, dimers, and trimers was optimized. The fitting results are shown in Figure 4, and a summary of the fit results is shown in Table 2. For the cage(5T), cage(6T), and cage(7T), the lowest q part of the data could not be reproduced by only monomers, dimers, and trimers. However, as the inclusion of higher-order oligomers would require some speculative assumptions of the structure of these, the data below $q = 0.018$ Å^{-1} were omitted from the fits. The results show that, similarly to the radius of gyration and maximum sizes, there is a systematic increase of the cage size R_{eq} with the linker length. Note that we define the cage size R_{eq} as the distance between the center of the cage and the center of each double DNA strand. For cage(3T), cage(4T), cage(5T), and cage(6T), the majority of the cages are monomeric with a small population of dimers and/or trimers. For cage(7T), a relatively large amount of trimers of the cages are present, as already presented previously.[7]

The size of the various oligomers of the cages can be compared with the sizes as determined by dynamic light scattering (DLS). For the obtained structure of the cages with different linker lengths, the hydrodynamic radii of the monomers, dimers, and trimers were calculated and compared with experimental data obtained from DLS experiments. A summary of the results is shown in Table 3. Since the DLS experiments were performed at different concentrations (and in some cases on different batches) of the cage samples, a direct comparison with the populations found by SAXS analysis is not meaningful. However, it is clear that there is in all cases a reasonable agreement between the theoretical

TABLE 2. Summary of the Fitting Results[a]

	3T	4T	5T	6T	7T
parameter	$\chi^2 = 7.5$	$\chi^2 = 4.4$	$\chi^2 = 7.7$	$\chi^2 = 7.5$	$\chi^2 = 7.7$
R_{eq} [Å]	50.0 ± 1.5	55.4 ± 0.3	55.8 ± 0.8	60 ± 1	62.5 ± 0.7
T_{angle} [°]	28 ± 4	24 ± 3	20 ± 3	24 ± 6	20 ± 3
R_{bead} [Å]	10 ± 2	13 ± 2	14 ± 2	12 ± 2	14 ± 2
d [Å]	124 ± 8	130 ± 13	130 ± 8	139 ± 9	142 ± 17
f_{mon}	0.68	0.91	0.90	0.72	0.44
f_{dim}	0.14	0.00	0.07	0.00	0.00
f_{trim}	0.18	0.09	0.03	0.28	0.56

[a]Summary of the fitting results shown in Figure 4, using the truncated octahedron model for the SAXS experimental data from samples with different thymidine linker lengths; χ^2 is the reduced chi-square value for the fits. R_{eq} denotes the cage size, as defined by the distance between the center of the cage and the center of each double-stranded arm. T_{angle} denotes the twist angle and is defined by the angle between the double-stranded arm and the symmetry plane of the octahedron. R_{bead}, d, f_{mon}, f_{dim}, and f_{trim} are parameters used in the IFT analysis.

TABLE 3. Summary of DLS results[a]

	3T			4T			5T			6T			7T		
	M	D	T	M	D	T	M	D	T	M	D	T	M	D	T
R_h Theo [Å]	61	83	102	64	89	112	65	87	108	68	93	114	71	99	124
	NW	MW		NW	MW		NW	MW		NW	MW		NW	MW	
R_h exp [Å]	57 ± 2	106 ± 10		73 ± 2	83 ± 2		68 ± 1	72 ± 1		77 ± 3	~113		76 ± 2	88 ± 3	

[a]Comparison between the theoretical (R_h Theo) hydrodynamic radius for the models obtained from the truncated octahedron model (monomers, M; dimers, D; and trimers, T) and the experimental values (R_h exp) obtained from DLS measurements for the samples with different linker lengths. NW denotes the number-averaged values, while MW denotes the mass-averaged values of the hydrodynamic radius.

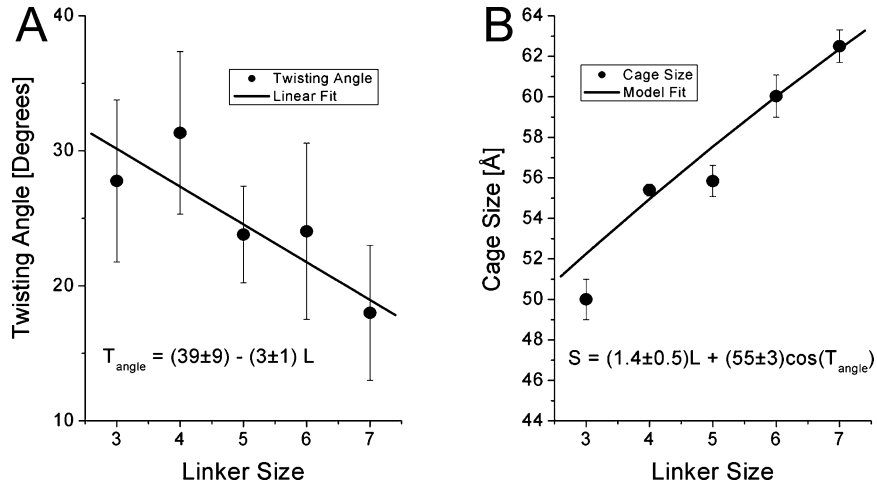


Figure 5. Variation of the twist angle and cage size as functions of the thymidine linker length. (A) Twist angle *vs* thymidine linker length *L*. A simple straight line was used to fit the data providing the expected twisting angle for a given thymidine linker length. (B) Cage size R_{eq} *vs* thymidine linker length *L*. The expression $R_{eq} = cL + d\cos(T_{angle})$ combining the linker length and the twisting angle was used for fitting the experimental points.

and experimental hydrodynamic radius. The number-averaged (NW) value provides information about the population containing the highest number of particles, and in this way, small populations of dimers and trimers do not affect the average hydrodynamic radius too much. The situation is different when analyzing the mass-averaged (MW) hydrodynamic radius value. In this case, even a small population of oligomers gives a strong contribution to the average hydrodynamic radius, and therefore, it is a good indicator for the presence of oligomers. Since for all samples the number-averaged (NW) and the mass-averaged (MW) hydrodynamic radii were always within the values from monomers and trimers, the DLS data also indicate the presence of oligomers in solution in agreement with the SAXS results.

Taken together, the SAXS analysis suggests that the monomeric cage(3T) to cage(6T) all have an overall 3D structure of a truncated octahedron similar to the structure of cage(7T) previously published.[7] The subset of oligomers present in each sample does not correlate with the linker length or any other properties of the cage in the samples and may, hence, best be explained as an artifact of the high concentration of the cage samples used for the SAXS and DLS analysis. This notion is supported by gel electrophoresis of the cage samples subjected to SAXS analysis, giving rise to only one distinct band with a mobility corresponding to monomeric cages (data not shown). At the more detailed level, from the least-squares fitting, we found the helix twist angle in the cage structures to increase as the linker length decreased, which may be interpreted as an increased constraint in structures with shortened linkers (Figure 5A). The twist angle is defined as the angle between the cage double-stranded helix and the symmetry plane of the truncated octahedron. Although the determination of the twist angle is not very accurate from the SAXS analysis (relatively large error bars), there is a clear trend, which can be described by a straight line: $T_{angle} = a + bL$, where *a* and *b* are fitting parameters and *L* is the linker length, which is equal to the number of thymidines. Also, a plot of the cage size R_{eq} *versus* the linker length *L* (Figure 5B) displays a clear trend. In this case, we described the variation of the cage size as a function of the linker length using the following expression: $R_{eq} = cL + d\cos(T_{angle})$, were *c* and *d* are fitting parameters and T_{angle} is the twist angle dependence obtained from the other plot. We choose this expression in order to take into account the projection of the DNA strand on the symmetry plane due to the twist angle. This simple description of the twisting angle and cage size as a function of the linker length indicates a correlated change of the twisting angle and cage size with the variation of the linker length.

Hydrogen Bond Deformation Energy of Cage(2T), Cage(3T), Cage(5T), and Cage(7T): Figure 6 shows the hydrogen bond deformation energy between each base pair in all of the 12 double-stranded arms forming the different truncated octahedral cages. In all cages, the base pairs located in the middle part of the double-stranded arms show optimal hydrogen bonds. In cage(3T), cage(5T), and cage(7T), the extremities of the arms are slightly deformed, showing average deformation energies of about 2.0 kcal/mol, while in the case of cage(2T), the average deformation energy is higher, being about 7.0 kcal/mol, with some peaks higher than 15.0 kcal/mol. This result indicates that the high local distortion felt by the double-stranded arms is the main reason for the lack of cage(2T) assembly. Such a high deformation is not present in the cage(3T), cage(5T), and cage(7T) models, providing an energetic explanation for the reasonably high experimental yields observed for these cages. Hence, the length of the single-stranded thymidine linker appears to be the crucial parameter for efficient assembly of 3D structures following the basic truncated octahedral cage design presented here with

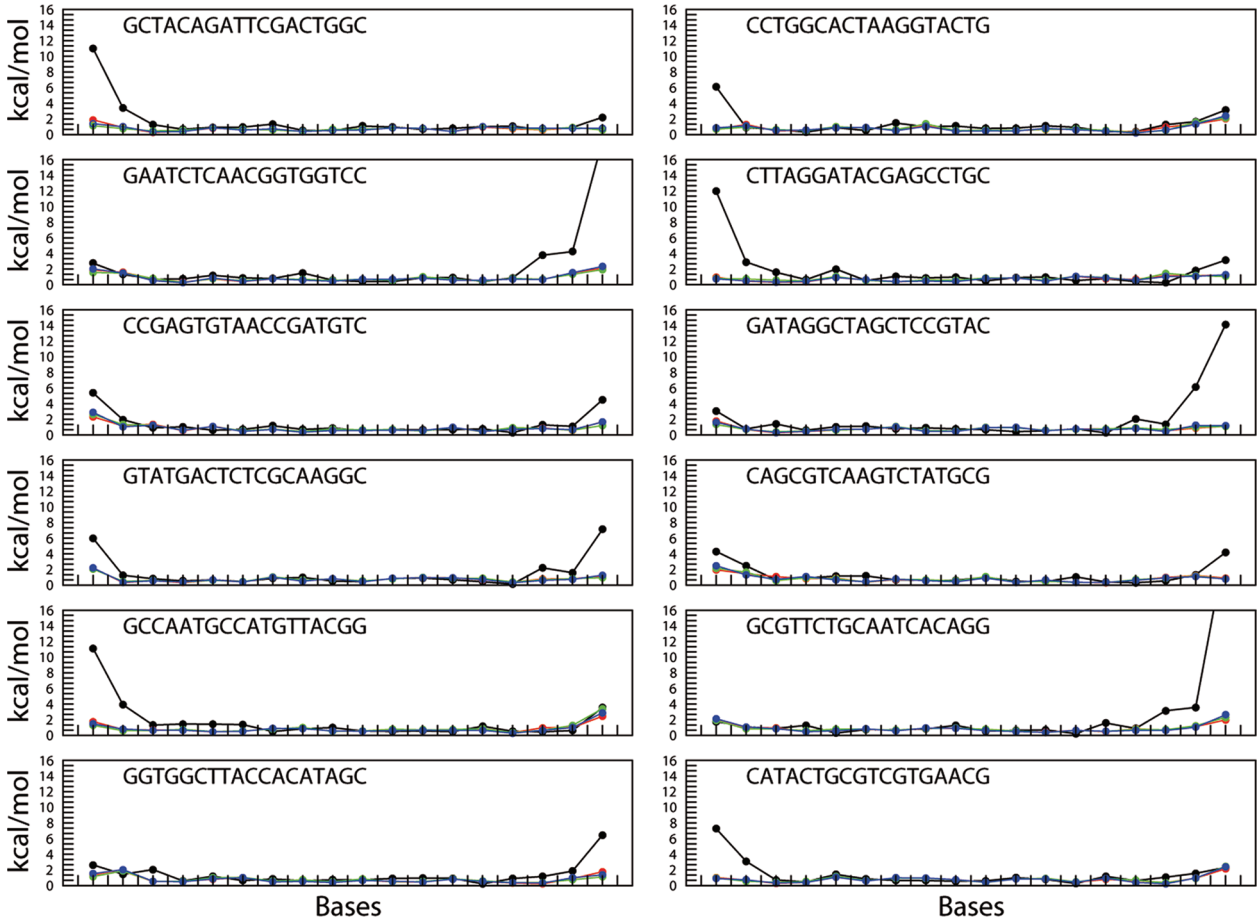


Figure 6. Hydrogen bond deformation energy for the different cages. The hydrogen bond deformation energy occurring in each base pair for all of the 12 double-stranded arms of the cage(2T) (black line), cage(3T) (red line), cage(5T) (blue line), and cage(7T) (green line) was calculated and plotted. The sequence of each double-stranded arm is shown at the top of one of the strands in each panel.

a minimum length of three nucleotides being required to obtain a cage and permit the right pairing of the base pairs in the directly connected double-stranded arms.

Taken together, the presented data reveal the important determinant for efficient assembly of the truncated octahedral DNA cage to be the length of the single-stranded linker regions in the structure. The presence of such regions imposes a certain degree of flexibility in the structure, which most likely explains the rather high one-step assembly yield of the cage structure compared to, for example, the cube and truncated octahedron, presented by Seeman and co-workers.[2,3] Although the presence of single-stranded regions introduces an inherent flexibility to the cage structure, the increased twist angle (between the double-stranded arms and the symmetry plane of the truncated octahedral cage) imposed by decreasing linker length suggests a more constrained structure and thereby probably more rigidity in cages with short linker regions. This possibility is currently under investigation.

CONCLUSION

In the present study, we demonstrated a direct correlation between the one-step assembly efficiency of a partly single-stranded truncated octahedral DNA cage and the length of the single-stranded regions of the structure. The presented investigations were based on a previously published structural framework (cage(7T)), which was built from 12 arms of double-stranded B-DNA interrupted by single-stranded linker regions of seven thymidines.[7] This structure assembled with a surprisingly high efficiency due to a rather high degree of flexibility imposed by the single-stranded regions, as suggested by MD simulations.[19] Consistently, we show in the present work that the linker length determines assembly of the cage structure, with three thymidines constituting the minimal linker length necessary to allow efficient cage assembly. The overall structures of DNA cages with successively decreasing linker regions from seven to three thymidines were identical except for their sizes, which correlated directly with the linker lengths. At the more detailed level, an increase in the angle between the double-stranded arms and the symmetry planes of the truncated octahedron cages with decreasing linker lengths suggests an increased constraint in the 3D structures when the linker regions are shortened. This, in turn, may indicate an increased rigidity of cages with shorter compared to longer linker regions, as previously suggested by MD simulations.[19] The inability of a cage structure with two thymidine linker

regions to assemble can be explained by energy minimization calculations demonstrating a highly increased deformation at the extremities of the double-stranded arms compared to cages with longer thymidine linkers. Taken together, the comprehensive analyses of the cage structures presented here may serve as useful guidelines for future design and efficient assembly of covalently closed 3D nanostructures composed of natural DNA conformations. In longer terms, such entities may serve as ideal units for the construction of large-order lattices potentially promoted *via* interaction or reaction with naturally occurring enzymes.

METHODS

Design of Oligonucleotides. The sequences of the oligonucleotides designed to assemble into cage(2T), cage(3T), cage(4T), cage(5T), cage(6T), and cage(7T) are shown Figure 1C. The sequences were chosen as described previously.[7]

Assembly of the DNA Cages. The partly single-stranded cage(2T), cage(3T), cage(4T), cage(5T), cage(6T), or cage(7T) was assembled by mixing the appropriate oligonucleotides (purchased from GeneLink) in equimolar amounts in an assembly buffer containing 70 mM Tris-HCl (pH 7.6), 10 mM $MgCl_2$, 5 mM dithiothreitol, and 1 mM ATP. Prior to assembly, phosphates were added to the 5′ ends of the oligonucleotides by incubation with T4 polynucleotide kinase (purchased from New England Biolabs) for 30 min at 37 °C. Subsequently, the samples were heated to 65 °C for 10 min and cooled by 0.25 °C/min until the sample reached a temperature of 30 °C. At this temperature, T4 DNA ligase (purchased from New England Biolabs) was added, and cooling was continued at 0.25 °C/min until 16 °C. To test covalent closure of fully assembled DNA cages, samples were heated to 95 °C for 5 min and quickly cooled to 4 °C.

Gel Electrophoresis and Purification of the DNA Cages. The partly and fully assembled cages were analyzed in native 5% polyacrylamide gels or purified from 5% denaturing gels as described previously.[7] For small-angle X-ray scattering (SAXS) analysis, an additional purification step was included after elution from the gel slice. The eluted cage was purified on a QIAGEN-tip 500 column (from QIAGEN Plasmid Maxi kit), essentially following the standard procedure from the kit. The cage was eluted from the column in 300 μL fractions, and EtOH precipitated. The fractions containing the highest concentrations of eight-stranded DNA cage (1−2 μg/μL) were used for SAXS analysis.

SAXS Measurements and Modeling. SAXS measurements were performed at the SAXS laboratory of the Department of Chemistry at the Aarhus University as described previously.[7,21] The SAXS intensity is recorded as a function of the scattering vector modulus $q = 4\pi \sin(\theta)/\lambda$, where λ is the radiation wavelength and 2θ is the scattering angle. The characteristic real-space distance distribution functions $p(r)$ were determined from the scattering data using indirect Fourier transformation[20] using the implementations described elsewhere.[22,23] This function corresponds to a histogram over all distances between pairs of points within the particle, and it gives direct insight into the particle shape and size. To model the DNA cage structure, we developed a method using rigid-body refinement and least-squares procedures.[7] Atomic coordinates of real DNA strands were used in the models for the cages. To speed up the calculations, only the C2* sequence was used to describe the DNA structure, which we have already shown to be a reasonable approximation.[7] The octahedral symmetry is applied to build the full structure, and the ends of consecutive paired DNA sides with 18 base pairs were connected using linear arrangements of spheres to mimic the thymidine linkers in this region. The scattering intensity is calculated using the Debye equation:[24]

$$P_{\text{model}}(q) = \frac{F_{\text{sphere}}(q, R_{\text{bead}})^2}{N^2} \sum_{i,j=1}^{N} \frac{\sin(qr_{ij})}{qr_{ij}} \quad (1)$$

where $F_{\text{sphere}}(q,R_{\text{bead}})$ is the form factor of a sphere of radius R_{bead} and r_{ij} is the distance between the centers of the ith and jth sphere. An approximation of this expression speeds up the calculation by building the histogram of distances $m(r_k)$:[25]

$$P_{\text{model}}(q) = \frac{F_{\text{sphere}}(q, R_{\text{bead}})^2}{A} \left(N + \sum_{i,j=1}^{N_{\text{bins}}} m(r_k) \frac{\sin(qr_k)}{qr_k} \right) \quad (2)$$

where A is a (trivial) normalization constant. The geometrical model is expressed in terms of a set of parameters (sizes, distances, and angles), and the form factor can be used in a least-squares procedure[26,27] which allows the optimization of the model structure by fitting to the experimental data. The analysis of the data showed presence of lower-order oligomers, and to obtain satisfactory fits to the experimental data, an intensity expression with combinations of dimers and trimers of the cage was used:

$$I(q) = S_{C_1} P(q)_{\text{model}} S(q)_{\text{total}} + S_{C_2} P_{\text{Deb}}(q) + \text{back} \quad (3)$$

Here S_{C_1} and S_{C_2} are scale factors for the overall intensity and Debye contribution and $P(q)_{\text{model}}$ is the averaged form factor calculated from eq 2. $S(q)_{\text{total}}$ is the total structure factor describing the presence of monomer, dimers, and trimers in solution and is given by

$$S(q)_{\text{total}} = f_{\text{mon}} + f_{\text{dim}} S(q)_{\text{dim}} + f_{\text{trim}} S(q)_{\text{trim}} \quad (4)$$

where f_{mon}, f_{dim}, and f_{trim} are the number fraction of the monomeric, dimeric, and trimeric cages, respectively, which fulfils: $\sum_i f = 1$. $S(q)_{\text{dim}}$ and $S(q)_{\text{trim}}$ were taken as the structure factors of linear arrangement of spherically symmetrical particles separated by a distance d:

$$S(q)_{\text{dim}} = 2 + 2\frac{\sin(qd)}{qd} \quad (5)$$

$$S(q)_{\text{trim}} = 3 + 4\frac{\sin(qd)}{qd} + 2\frac{\sin(2qd)}{2qd} \quad (6)$$

$P_{\text{Deb}}(q)$ is the Debye form factor expression for Gaussian chains given by[28]

$$P_{\text{Deb}}(q, R_g) = 2[\exp(-u) + u - 1]/u^2,\ u = R_g^2 q^2 \quad (7)$$

The Gaussian chain scattering is introduced in order to account for possible misfolded strands that give a polymer-like scattering contribution. The addition of this term is essential to describe the scattering at high scattering vectors.

Dynamic Light Scattering: The experiments were performed at the light scattering laboratory of the Department of Chemistry at Aarhus University. The measurements were performed at a scattering angle of 90 °C on an ALV instrument (Langen, Germany) with a CGS-8F goniometer system equipped with an ALV-6010/EPP multitau digital correlator and a helium−neon diode laser (JDS Uniphase). ALV software was used for deriving the hydrodynamic radius using cumulant analysis and inverse Laplace transforms. The samples were similar to the ones used in the SAXS experiments but were measured at 23 °C. The hydrodynamic radii of the monomer, dimer, and trimer models as obtained by SAXS were calculated using program HYDROPRO.[29]

Energy Minimization and Hydrogen Bond Deformation Energy Calculations. The model was built and optimized in two stages. First, a coarse-grained optimization was done in which the dimensions of the cages obtained from SAXS were employed together with the chemical structure of the linkers. Subsequently, an energy minimization was carried out, and the hydrogen

bonding energies along the helices were estimated. In the coarse-grained optimization, the 12 helices forming the cages were modeled through the program 3DNA[30] and assembled to obtain the complete cage structure through a Python in-house written program. The program uses the biopython module[31] to arrange, in the ideal truncated octahedral geometry, the prebuilt double-stranded DNA helices that are treated as independent rigid bodies. The optimal helix orientation and relative distance have been obtained in the absence of the single-stranded thymidine linkers, imposing a geometrical constraint, minimizing the following formula:

$$S = \sum_{j=1}^{12} k(R_j - R_{eq})^2 + \sum_{i=1}^{24} (D_i - D_{eq})^2 \qquad (8)$$

where S represent the score indicating the displacement of the DNA helices from their ideal position; j is the helix index; R_j is the distance of the geometric center of the DNA helix from the geometric center of the cage; R_{eq} is the size of the cage, that for the cage(2T), cage(3T), cage(5T), and cage(7T) is 48, 50, 56, and 63 Å, respectively; k is a constant of value 500 that has been used to avoid an excessive cage radius variation; i is the thymidine linker index; D_i is the length of each thymidine linker, and D_{eq} represents the average distance between the fully stretched and the B-form packed conformations of the thymidine linker (D_{eq}(2T) = 13 Å, D_{eq}(3T) = 17 Å, D_{eq}(5T) = 28 Å, D_{eq}(7T) = 40 Å).

For cage(3T), cage(5T), and cage(7T), the R_{eq} values were obtained from the SAXS experiment (see Table 3), confirmed in cage(7T) by molecular dynamics simulation.[19] In cage(2T), the R_{eq} value was imposed to maintain a distance between the DNA helices spanned by a two thymidines linker. The minimization was carried out using the Powell algorithm[32] embedded in the python *scipy* module (http://www.scipy.org). Once the optimal DNA helix orientations had been reached, the thymidine linkers were manually added using the program SYBYL (TRIPOS, http://www.tripos.com/), and the clashes were removed through the SYBYL anneal module. Subsequently, each cage structure was regularized through 5000 steps of energy minimization (salt concentration of 20 mM and electrostatic cutoff of 12 Å) using the generalized Born solvent model[33,34] implemented in the program AMBER 9.0[35] using the AMBER03 force field.[36,37] For cage(2T), cage(3T), cage(5T), and cage (7T), the energy penalty due to deformations in the DNA helices induced by the minimization algorithm was calculated through the program deform_energy (http://3dna.rutgers.edu) that evaluates the hydrogen bond deformation in each base pair using parameters derived from MD simulations.[38]

Acknowledgment. The study was supported by The Lundbeck Foundation, The Danish Cancer Society, The Danish Council for Independent Research, Natural Sciences (FNU), The Novo Nordisk Foundation, Aase og Ejnar Danielsen's Foundation, Arvid Nilsson's Foundation, The Augustinus Foundation, The Beckett Foundation, Brødrene Hartmann's Foundation, Civilingeniør Frode V. Nyegaard og hustrus Foundation, Direktør Einar Hansen og hustru fru Vera Hansen's Foundation, Fabrikant Einar Willumsen's Mindelegat, Fru Astrid Thaysen's Legat, The Harboe foundation, Karen Elise Jensen's Foundation, Kong Christian den Tiendes Foundation, Købmand Sven Hansen og hustru Ina Hansen's Foundation, Krista og Viggo Petersen's Foundation.

REFERENCES AND NOTES

1. Seeman, N. C. An Overview of Structural DNA Nanotechnology. *Mol Biotechnol.* **2007**, *37*, 246–257.
2. Chen, J. H.; Seeman, N. C. Synthesis from DNA of a Molecule with the Connectivity of a Cube. *Nature* **1991**, *350*, 631–633.
3. Zhang, Y.; Seeman, N. C. The Construction of a DNA Truncated Octahedron. *J. Am. Chem. Soc.* **1994**, *116*, 1661–1669.
4. Shih, W. M.; Quispe, J. D.; Joyce, G. F. A 1.7-kilobase Single-Stranded DNA That Folds into a Nanoscale Octahedron. *Nature* **2004**, *427*, 618–621.
5. Aldaye, F. A.; Sleiman, H. F. Modular Access to Structurally Switchable 3D Discrete DNA Assemblies. *J. Am. Chem. Soc.* **2007**, *129*, 13376–13377.
6. Goodman, R. P.; Schaap, A. T.; Tardin, C. F.; Erben, C. M.; Berry, R. M.; Schmidt, C. F.; Turberfield, A. J. Rapid Chiral Assembly of Rigid DNA Building Blocks for Molecular Nanofabrication. *Science* **2005**, *310*, 1661–1665.
7. Andersen, F. F.; Knudsen, B.; Oliveira, C. L. P.; Frøhlich, R. F.; Krüger, D.; Bungert, J.; Agbandje-McKenna, M.; McKenna, R.; Juul, S.; Koch, J.; *et al.* High-Yield Assembly of a Stable Nanoscale DNA Cage. *Nucleic Acids Res.* **2008**, *36*, 1113–1119.
8. He, Y.; Ye, T.; Su, M.; Zhang, C.; Ribbe, A. E.; Jiang, W.; Mao, C. Hierarchical Self-Assembly of DNA into Symmetric Supramolecular Polyhedra. *Nature* **2008**, *452*, 198–201.
9. Simmel, F. C. Three-Dimensional Nanoconstruction with DNA. *Angew. Chem., Int. Ed.* **2008**, *47*, 5884–5887.
10. Andersen, E. S.; Dong, M.; Nielsen, M. M.; Jahn, K.; Subramani, R.; Mamdouh, W.; Golas, M. M.; Sander, B.; Stark, H.; Oliveira, C. L.; *et al.* Self-Assembly of a Nanoscale DNA Box with a Controllable Lid. *Nature* **2009**, *459*, 73–76.
11. Endo, M.; Sugiyama, H. Three-Dimensional DNA Nanostructures Constructed by Folding of Multiple Rectangles. *Nucleic Acids Symp. Ser.* **2009**, *53*, 81–82.
12. Li, Z.; Wei, B.; Nangreave, J.; Lin, C.; Liu, Y.; Mi, Y.; Yan, H. A Replicable Tetrahedral Nanostructure Self-Assembled from a Single DNA Strand. *J. Am. Chem. Soc.* **2009**, *131*, 13093–13098.
13. Bhatia, D.; Mehtab, S.; Krishnan, R.; Indi, S. S.; Basu, A.; Krishnan, Y. Icosahedral DNA Nanocapsules by Modular Assembly. *Angew. Chem., Int. Ed.* **2009**, *48*, 4134–4137.
14. Kato, T.; Goodman, R. P.; Erben, C. M.; Turberfield, A. J.; Namba, K. High-Resolution Structural Analysis of a DNA Nanostructure by CryoEM. *Nano Lett.* **2009**, *9*, 2747–2750.
15. Aldaye, F. A.; Lo, P. K.; Karam, P.; McLaughlin, C. K.; Cosa, G.; Sleiman, H. F. Modular Construction of DNA Nanotubes of Tunable Geometry and Single- or Double-Stranded Character. *Nat. Nanotechnol.* **2009**, *4*, 349–352.
16. Hede, M. S.; Petersen, R. L.; Frøhlich, R. F.; Krüger, D.; Andersen, F. F.; Andersen, A. H.; Knudsen, B. R. Resolution of Holliday Junction Substrates by Human Topoisomerase I. *J. Mol. Biol.* **2007**, *365*, 1076–1092.
17. Andersen, F. F.; Andersen, K. E.; Kusk, M.; Frøhlich, R. F.; Westergaard, O.; Andersen, A. H.; Knudsen, B. R. Recombinogenic Flap Ligation Mediated by Human Topoisomerase I. *J. Mol. Biol.* **2003**, *330*, 235–246.
18. Knudsen, B. R.; Lee, J.; Lisby, M.; Westergaard, O.; Jayaram, M. Alcoholysis and Strand Joining by the Flp Site-Specific Recombinase. Mechanistically Equivalent Reactions Mediated by Distinct Catalytic Configurations. *J. Biol. Chem.* **1998**, *273*, 22028–22036.
19. Falconi, M.; Oteri, F.; Chillemi, G.; Andersen, F. F.; Tordrup, D.; Oliveira, C. L.; Pedersen, J. S.; Knudsen, B. R.; Desideri, A. Deciphering the Structural Properties That Confer Stability to a DNA Nanocage. *ACS Nano* **2009**, *3*, 1813–1822.
20. Glatter, O. A New Method for the Evaluation of Small-Angle Scattering Data. *J. Appl. Crystallogr.* **1977**, *10*, 415–421.
21. Pedersen, J. S. A Flux- and Background-Optimized Version of the NanoSTAR Small-Angle X-ray Scattering Camera for Solution Scattering. *J. Appl. Crystallogr.* **2004**, *37*, 369–380.
22. Pedersen, J. S.; Hansen, S.; Bauer, R. The Aggregation Behavior of Zinc-Free Insulin Studied by Small-Angle Neutron Scattering. *Eur. Biophys. J.* **1994**, *22*, 379–389.
23. Oliveira, C. L. P.; Behrens, M. A.; Pedersen, J. S.; Erlacher, K.; Otzen, D.; Pedersen, J. S. A SAXS Study of Glucagon Fibrillation. *J. Mol. Biol.* **2009**, *387*, 147–161.
24. Debye, P. Zerstreuung von Röntgenstrahlen. *Ann. Phys. Leipzig* **1915**, *46*, 809–823.
25. Glatter, O. Kratky, O. *Small Angle X-ray Scattering*; Academic Press: London, 1982; Chapter 4, pp 157−162.
26. Pedersen, J. S. Small-Angle Scattering from Precipitates: Analysis by Use of a Polydisperse Hard-Sphere Model. *Adv. Colloid Interface Sci.* **1997**, *70*, 171–210.

27. Pedersen, J. S. Instrumentation for Small-Angle X-ray and Neutron Scattering and Instrumental Smearing Effects. In *Neutrons, X-Rays and Light: Scattering Methods Applied to Soft Condensed Matter*; North-Holland: Amsterdam, 2002; Chapter 16, pp 127−144.
28. Debye, P. Molecular-Weight Determination by Light Scattering. *J. Phys. Colloid Chem.* **1947**, *51*, 18–32.
29. Garcia De La Torre, J.; Huertas, M. L.; Carrasco, B. Calculation of Hydrodynamic Properties of Globular Proteins from Their Atomic-Level Structure. *Biophys. J.* **2000**, *78*, 719–730.
30. Lu, X.-J.; Olson, W. K. 3DNA: A Software Package for the Analysis, Rebuilding and Visualization of Three-Dimensional Nucleic Acid Structures. *Nucleic Acids Res.* **2003**, *31*, 5108–5121.
31. Cock, P. J. A.; Antao, T.; Chang, J. T.; Chapman, B. A.; Cox, C. J.; Dalke, A.; Friedberg, I.; Hamelryck, T.; Kauff, F.; Wilczynski, B.; *et al.* Biopython: Freely Available Python Tools for Computational Molecular Biology and Bioinformatics. *Bioinformatics* **2009**, *25*, 1422–1423.
32. Powell, M. J. D. An Efficient Method for Finding the Minimum of a Function of Several Variables without Calculating Derivatives. *Comput. J.* **1964**, *17*, 155–162.
33. Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Pairwise Solute Descreening of Solute Charges from a Dielectric Medium. *Chem. Phys. Lett.* **1995**, *246*, 122–129.
34. Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Parametrized Models of Aqueous Free Energies of Solvation Based on Pairwise Descreening of Solute Atomic Charges from a Dielectric Medium. *J. Phys. Chem.* **1996**, *100*, 19824–19839.
35. Case, D. A.; Cheatham, T. E., III; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A., Jr.; Simmerling, C.; Wang, B.; Woods, R. The AMBER Biomolecular Simulation Programs. *J. Comput. Chem.* **2005**, *26*, 1668–1688.
36. Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T. A Point-Charge Force Field for Molecular Mechanics Simulations of Proteins Based on Condensed-Phase Quantum Mechanical Calculations. *J. Comput. Chem.* **2003**, *24*, 1999–2012.
37. Ponder, J. W.; Case, D. A. Force Fields for Protein Simulations. *Adv. Protein Chem.* **2003**, *66*, 27–85.
38. Lankas, F.; Sponer, J.; Langowski, J.; Cheatham, T. E. DNA Basepair Step Deformability Inferred from Molecular Dynamics Simulations. *Biophys. J.* **2003**, *85*, 2872–2883.