



## General Strategy for the Design of DNA Coding Sequences Applied to Nanoparticle Assembly

Théo Calais, Vincent Baijot, Mehdi Djafari Rouhani, David Gauchard, Yves J Chabal, Carole Rossi, Alain Estève

### ► To cite this version:

Théo Calais, Vincent Baijot, Mehdi Djafari Rouhani, David Gauchard, Yves J Chabal, et al.. General Strategy for the Design of DNA Coding Sequences Applied to Nanoparticle Assembly. *Langmuir*, 2016, 32 (37), pp.9676-9686. 10.1021/acs.langmuir.6b02843 . hal-01368340

**HAL Id: hal-01368340**

**<https://hal.science/hal-01368340>**

Submitted on 19 Jun 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# General Strategy for the Design of DNA Coding Sequences Applied to Nanoparticle Assembly

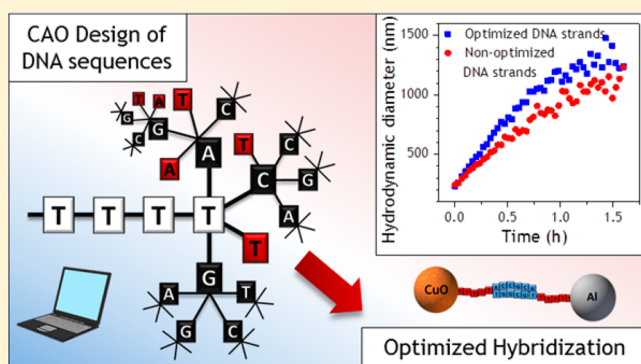
Théo Calais,<sup>†</sup> Vincent Bajiot,<sup>†</sup> Mehdi Djafari Rouhani,<sup>†</sup> David Gauchard,<sup>†</sup> Yves J. Chabal,<sup>‡</sup> Carole Rossi,<sup>\*,†</sup> and Alain Estève<sup>†</sup>

<sup>†</sup>LAAS-CNRS, Université de Toulouse, CNRS, Toulouse, France

<sup>‡</sup>Department of Materials Science and Engineering, University of Texas at Dallas, Richardson, Texas 75080, United States

## S Supporting Information

**ABSTRACT:** The DNA-directed assembly of nano-objects has been the subject of many recent studies as a means to construct advanced nanomaterial architectures. Although much experimental *in silico* work has been presented and discussed, there has been no in-depth consideration of the proper design of single-strand sticky termination of DNA sequences, noted as ssST, which is important in avoiding self-folding within one DNA strand, unwanted strand-to-strand interaction, and mismatching. In this work, a new comprehensive and computationally efficient optimization algorithm is presented for the construction of all possible DNA sequences that specifically prevents these issues. This optimization procedure is also effective when a spacer section is used, typically repeated sequences of thymine or adenine placed between the ssST and the nano-object, to address the most conventional experimental protocols. We systematically discuss the fundamental statistics of DNA sequences considering complementarities limited to two (or three) adjacent pairs to avoid self-folding and hybridization of identical strands due to unwanted complements and mismatching. The optimized DNA sequences can reach maximum lengths of 9 to 34 bases depending on the level of applied constraints. The thermodynamic properties of the allowed sequences are used to develop a ranking for each design. For instance, we show that the maximum melting temperature saturates with 14 bases under typical solvation and concentration conditions. Thus, DNA ssST with optimized sequences are developed for segments ranging from 4 to 40 bases, providing a very useful guide for all technological protocols. An experimental test is presented and discussed using the aggregation of Al and CuO nanoparticles and is shown to validate and illustrate the importance of the proposed DNA coding sequence optimization.



## INTRODUCTION

The interest in DNA nanotechnology to program the assembly of nanoparticles into macroscopic nanocomposites emerged in the 1990s.<sup>1,2</sup> Undoubtedly, the controlled interplay of DNA complementary and noncomplementary strands made DNA nanotechnologies one of the most powerful bottom-up approaches to building hierarchical architectures of nano-objects (noble metals, semiconductors, oxides, and polymers) leading to an almost infinite variety of high-performance programmable DNA/nanoparticles hybrid materials. One mainstream DNA-based assembly approach consists of directing the assembly of colloids of interest, mostly gold nanoparticles, by taking advantage of the thiol/metal chemistry<sup>3</sup> to covalently attach DNA strands to nanoparticle surfaces. Other chemical alternatives have also been investigated, such as antigen/antibody-like binding.<sup>4,5</sup> Since the seminal work by Alavistas and Mirkin<sup>1,2</sup> on gold nanoparticles, many DNA/nanoparticle assembly processes have been reported, notably by varying the DNA length and processing conditions and with a consideration of other materials for

applications in catalysis,<sup>6</sup> spectroscopy,<sup>7–10</sup> optical devices,<sup>11,12</sup> and energetic materials,<sup>13</sup> to name just a few.

In practice, the overall optimization process for controlling DNA-directed assembly in the liquid phase is a very complex task. Over the last several decades, substantial progress has been achieved in understanding the wet chemical technique parameters for the patterning of matter at the nanoscale, especially the effect of salt concentration on both the grafting density and hybridization of DNA strands<sup>5,14,15</sup> or the heterogeneity of DNA grafted on anisotropic nanoparticles as demonstrated very recently.<sup>16</sup> Highly ordered fcc and bcc structures of 5 or 10 nm gold nanoparticles, with different DNA strand lengths and a programmable interparticle distance<sup>17–21</sup> and more recently with the crystallization of nanoparticles into Wulff polyhedra<sup>22</sup> or anisotropic nanoparticles into more complex structures,<sup>23,24</sup> have been reported.

Received: July 31, 2016

Revised: August 29, 2016

However, despite great advances in the understanding of these DNA/nanoparticles composite systems, there are still numerous open questions, such as the relative number of DNA connections formed during assembly, the thermodynamics and kinetics of DNA binding events as a function of programmable oligonucleotide interactions, and the lengths that are responsible for particle assembly.

On the theoretical side, a number of approaches, based on either analytical<sup>20,25</sup> or numerical models,<sup>26–34</sup> have been elaborated to understand and predict the interaction and crystallization of self-assembled nanoparticles as a function of kinetics,<sup>28</sup> nanoparticle shape,<sup>29</sup> electrostatics,<sup>30,31</sup> and nanoparticle stoichiometry,<sup>32,33</sup> or the single-strand sticky termination of DNA sequences.<sup>34</sup> For example, Macfarlane et al. performed molecular dynamics simulations,<sup>33</sup> using a coarse-grained model, to investigate the crystallization of 15 nm gold nanoparticles and obtained results consistent with experimental observations. In these simulations, a double-stranded DNA, with different lengths, is used as the spacer. The sticky-end recognition part is short and constituted of only four single-stranded bases (GCGC). Such an empirical approach would no longer be adequate for longer oligonucleotide sequences. Although the length of the spacer part has been investigated,<sup>5</sup> the choice of the oligonucleotide sequence follows a priori an empirical design and has not been widely studied nor commented fundamentally. Seifpour et al. used a coarse-grained model to determine the influence of the percentage and localization of G/C bases on the DNA sequence on the assembly thermodynamics.<sup>34</sup> However, this model did not consider electrostatics and nonspecific interactions of bases with the nanoparticle surface. A rational design method has also been proposed.<sup>35</sup> The method is based on local constraints imposed on the DNA strands, with the goal of designing sequences showing a narrow distribution of melting temperatures. Within the nearest-neighbor approximation of thermodynamic properties, the constraints are limited to two adjacent bases. The authors use Euclidean graph adjacency matrices to find all possible sequences with a predefined melting temperature. Very recently, a test tube design method, based on thermodynamic equilibrium considerations, has been elaborated.<sup>36</sup> The authors design DNA sequences that would normally lead to target structures when mixed in a test tube at specified concentrations. Various innovative algorithms allow for the simplification of the calculation of partition functions, used to determine thermodynamic properties of the mixture at equilibrium. The starting sequences are chosen at random and are progressively improved by iterative mutations. Herein, we present a methodology to design DNA strands imposing constraints that are not only local but also take into account the long-range correlations of local sequences. These sequences, at high concentrations, can also be used as input for test tube design. Our algorithm reproduces all DNA sequences of any length, fulfilling the imposed constraints. The length of the strands is naturally limited by the constraints on long-range correlations.

In practice, we design optimized DNA sticky ends with coding lengths from 1 to 10 nm, fulfilling the most conventional requirements of DNA sticky ends encountered in the literature. Sticky ends usually refer to DNA single-stranded terminations extending from a double-stranded region and available for further hybridization. In the present article, the coding sequence region that is designed to interact with complementary strands follows a DNA single-strand spacer that

is supposed to be neutral during hybridization and overall assembly processes. In what follows, we use the term single-strand sticky termination noted as ssST. We also experimentally demonstrate the importance of the design of the DNA sequences in the case of the self-assembly of Al and CuO nanoparticles by comparing the assembly kinetics of nanoparticles functionalized with DNA strand sequences designed by our tool with sequences proposed and designed in 2013 by Gang et al.<sup>21</sup>

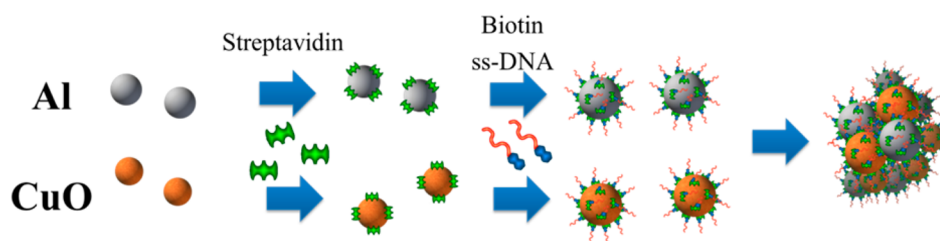
The results clearly demonstrate that DNA strand optimization by the new algorithm leads to better control of the assembly process, allowing increased thermal stability, aggregation kinetics, and improved composition. We specifically demonstrate that optimizing ssST to get double-stranded DNA with the highest melting temperature enables better control of the hybridization with temperature. This methodology represents a major advance for DNA-based nanoparticle assembly and materials engineering because it effectively improves the control of the hybridization efficiency, the double-stranded DNA melting temperature, and consequently the assembly reversibility.

The article is organized as follows. The first section presents the methodology and computational details. Then, we discuss the statistics of reported sequences as a function of the imposed constraints in terms of the degree of pairing complementarity and the presence or absence of DNA-based spacers (i.e., a series of thymines). The allowed sequences are then sorted by their melting temperatures. Finally, the method is tested by an experiment in which the thermodynamics and kinetics of the self-assembly of Al and CuO nanoparticles into microscopic nanocomposites are studied by diffraction light scattering (DLS) and scanning electron microscopy (SEM) using two types of DNA strands: one set of complementary DNA strands with ssST designed by our method and one set of DNA strands with ssST proposed earlier in the literature.

## EXPERIMENTAL SECTION

**Materials.** NaCl powder, monobasic and dibasic sodium  $\text{NaH}_2\text{PO}_4$  and  $\text{NaHPO}_4$  powders, phosphate-buffered saline 10X, and surfactant Tween 20 were purchased from Sigma-Aldrich. Phosphate buffer (0.2 M) at pH 7 was obtained by mixing monobasic and dibasic powders (39 and 61 vol %, respectively). NaCl solutions (2 M) were prepared by the dissolution of NaCl powder in ultrapure water. Streptavidin protein was purchased from Sigma-Aldrich. Dried biotin-modified oligonucleotides were purchased from Eurogentec. As soon as received, oligonucleotides were diluted in ultrapure water in stock solution concentrated to 0.5 mM and stored away from light at  $-20^\circ\text{C}$ . Before use, solutions were warmed to room temperature and homogenized by vortex mixing. Copper oxide (CuO, 50 nm) nanopowders were purchased from Sigma-Aldrich. Aluminum (Al, 70 nm) nanopowders were purchased from US Research Nanomaterials (Austin, TX). Diameters were confirmed by SEM imaging (Supporting Information SI-1 and SI-2).

**Methods.** Sonication was performed using a VibraCell VCX 500 ultrasonic probe system at 200 W, with 2 s pulses separated by 1 s. The temperature was maintained below  $40^\circ\text{C}$  in order to avoid the oxidation of Al nanoparticles (NPs). A Zetasizer Nano ZS instrument (Malvern Instruments) was used to determine the NP hydrodynamic diameter by diffraction light scattering (DLS), and the zeta potential was determined by Doppler laser electrophoresis. The signal of a He/Ne laser emitting at 632.8 nm traversed a measurement cell containing 150 and 850  $\mu\text{L}$  of a colloidal suspension for hydrodynamic and zeta potential analyses, respectively. The scattering intensity was measured at  $173^\circ$  with respect to the source. All of the zeta potential and hydrodynamic diameter measurements were performed at  $25^\circ\text{C}$  and at pH 7.3. A module added to the apparatus allowed us to control the



**Figure 1.** Schematic of the functionalization of Al and CuO nanoparticles. After dispersion in aqueous solvent (phosphate buffer (PB) 10 mM and Tween 0.05 vol %), nanoparticles are first functionalized with streptavidin. After rinsing in phosphate-buffered saline 0.1× with Tween 20 at 0.05 vol %, biotin-modified DNA strands are added to solution. After being rinsed, colloidal suspensions of functionalized Al and CuO NPs are dispersed in PBS 0.1× and Tween 0.05% vol. Finally, solutions are mixed in order to induce the self-assembly of nanoparticles.

temperature of the analysis from 10 to 90 °C. SEM imaging was performed on a Hitachi S4800 instrument. The samples were prepared by depositing a droplet of colloidal suspension on a TEM carbon-coated nickel grid.

The functionalization process of Al and CuO nanoparticles with DNA is based on a streptavidin–biotin complex strategy and includes four steps, as previously described<sup>13</sup> and illustrated in Figure 1:

(1) Suspension of Al and CuO nanoparticles in aqueous solution composed of 10 mM phosphate buffer (in order to stabilize the pH at 7, limiting nanoparticle oxidation<sup>37</sup>), 0.05 vol % Tween 20 surfactant to stabilize nanoparticles, and ultrapure deionized water.

(2) Streptavidin grafting of Al and CuO nanoparticles in an aqueous solution of 10 mM PB and 0.05 vol % Tween. After incubation overnight, solutions were rinsed three times with PBS 0.1× and Tween 20.

(3) Grafting of biotin-modified single-stranded DNA on Al and CuO nanoparticles in an aqueous solution of PBS 0.1× and 0.05% Tween 20. After overnight incubation, solutions were rinsed three times with PBS 0.1× and 0.05% Tween 20. The average diameter of DNA-functionalized nanoparticles is 230 nm for both Al and CuO nanoparticles, as determined by DLS measurements.

(4) Mixing of the two colloidal suspensions in an equimolar volume. An addition of NaCl allows the aggregation of nanoparticles by DNA self-assembly.

Supporting Information files SI-1 to SI-3 provide SEM images and DLS and zeta potential data for each step in the DNA functionalization process for Al and CuO nanoparticles. Initially and before functionalization, the suspension contains small aggregates of a few nanoparticles. Then, sonication breaks aggregates and leads to a significant reduction of the hydrodynamic diameter to 185 and 220 nm for CuO and Al, respectively. After functionalization, the increase in the average hydrodynamic diameters is 20 and 35 nm for Al and CuO nanoparticles, respectively. This gain corresponds to the size of oligonucleotides (30 bases long) plus the proteins grafted onto the nanoparticles. Note that the larger diameter measured for CuO nanoparticles is mostly due to the heterogeneous shape of nanoparticles.

The concentration of NPs, measured using atomic absorption spectroscopy after the dissolution of NPs, was estimated to be  $1.2 \times 10^{10}$  and  $5.7 \times 10^9$  NPs/mL for Al and CuO NPs, respectively. Grafting densities of DNA were determined by a fluorescence method and estimated to be  $3 \times 10^4$  single strands of DNA per nanoparticle, being 40 and 50 pmol/cm<sup>2</sup> for Al and CuO nanoparticles, respectively. The hybridization efficiency is evaluated to be 25 to 60%. A complete description of the characterization method, optimization process, and results will be presented in a future article.

For all kinetics analyses, 500  $\mu$ L of a mixture of Al and CuO NPs functionalized with DNA was prepared as follows: 440  $\mu$ L of Al and CuO colloidal suspensions was mixed with an equimolar ratio of both NPs, and then 60  $\mu$ L of solvent was added to the solution with a variable salt concentration (to set the nanoparticles concentration independently of the salt concentration). After homogenization, 150  $\mu$ L was removed for DLS measurements.

The aggregation kinetics of Al and CuO NPs were monitored in real time using DLS. The intensity autocorrelation functions were analyzed

by a general purpose algorithm integrated into the Malvern Zetasizer software in order to determine the mean hydrodynamic diameter and the distribution size of nanoparticles in solution. In this way, we can follow the evolution of the mean hydrodynamic diameter of the composite structures made of agglomerated NPs in solution as a function of time and temperature. The role of the sequence of DNA strands can be assayed by comparing the assembly kinetics of Al and CuO NPs functionalized with DNA strands from Table 2.

**Design of DNA Coding Sequences.** When no constraint is imposed, the number of different DNA strands of a given length  $N$  amounts to  $4^N$ . This number becomes prohibitive very soon, with increasing  $N$ , for an individual evaluation against any imposed constraint. This means that we cannot first construct the whole set of strands and then proceed to their individual evaluation. The only possible way is to restrict the evaluation process to a very small subset of the strands. The size of this subset will strongly depend on the complexity of the evaluation, whether thermodynamic, kinetic, or just topological. If we present the growing strand by a growing tree, then each node is separated into four branches corresponding to the four A, C, G, and T bases. To end up with a reasonable number of branches, the undesired branches of the growing tree should be cut at each node. This is the reason that we propose a bottom-up approach, where DNA strands are constructed one base at a time, tested, and immediately discarded if this base does not fulfill the required constraints. Topological constraints, including their long-range correlations, are particularly adapted to this approach that allows a reduction of the number of possible strands by  $\sim 10^{-14}$ . On the other hand, thermodynamic properties, such as the melting temperature, need a global evaluation of the individual strands over their small size subset.

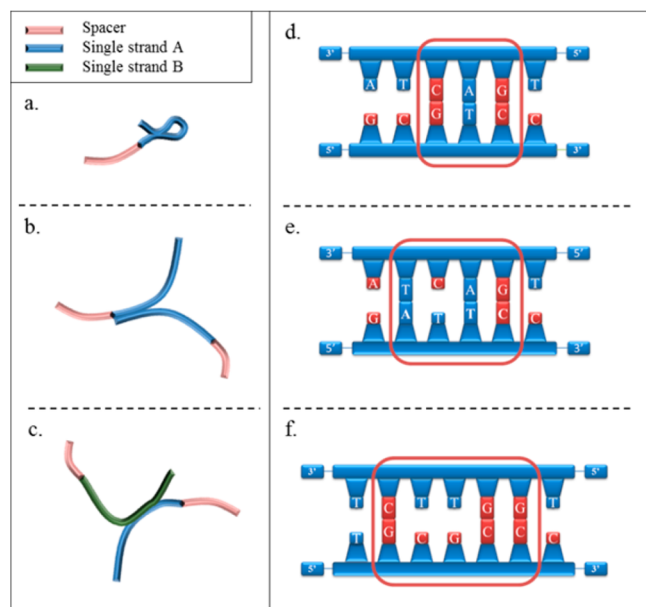
To implement this bottom-up approach, we introduce the concept of a compatibility tensor. (See also Supporting Information SI-4 for further details and the general algorithm.)

Each of the constraints is represented by a particular tensor, with its rank depending on the nature and the complexity of the imposed constraint. The tensor elements are of a binary nature: 0 for allowed sequences and 1 for forbidden ones.

Extending a given allowed sequence of length  $N$  to  $N + 1$ , we build different compatibility tensors associated with each constraint by going through the sequence and inverting initial 0 elements to 1. At the end of this process, the remaining 0 elements in the tensors indicate the new bases allowed for an extension of the sequence; in other words, 1 elements specify what branches should be cut at this stage. In the Results and Discussion section, this process is detailed for the construction of the first bases of the strands, in the case of the simplest constraints. In the Supporting Information, we give an example of how to extend longer strands with more sophisticated constraints. The use of compatibility tensors enables us to synthesize rapidly all topological elements present in the sequence, with respect to each imposed constraint. As we progress in the construction of a given DNA strand, the number of forbidden sequences increases more or less rapidly, depending on the imposed constraints, and tensor elements are gradually changed to 1. When one of the compatibility tensors is completed, the sequencing process is automatically stopped for the given sequence. At some point, all of the sequences reach their

maximal lengths, depending again on the imposed constraints. There is no other sequence beyond this point that can obey the constraints.

We consider four types of constraints illustrated and described in Figure 2: strand folding (SF), strand to strand (STS), partial



**Figure 2.** Schematic of the constraints used in the program. Oligonucleotides are represented by bars: red bars correspond to spacers (as, for example, repeated thymine bases), blue bars correspond to the sequence designed by the algorithm, and green bars correspond to the complementary strand of the designed oligonucleotide. (a) Avoided self-interaction of a single strand through the folding process, (b) hindered interactions (total or partial hybridization) between two strands with the same sequences, and (c) undesired partial hybridization of sticky-end sequences in the presence of their complementary counterpart. (d, e) Detailed and local sequence constraint: (d) a maximum of three adjacent bases are allowed to hybridize and (e, f) mismatched sequences (noted as M) that are forbidden when considering interactions.

hybridization (PH), and mismatch (MM). Figure 2a shows undesired SF, to be avoided. Folding happens when complementary bases are present within the strand. The minimum length of the loop allowing the folding is considered to be an input parameter in our algorithm. This parameter is set to four bases by default and in all applications reported in this article. Similarly, undesired STS interactions are represented in Figure 2b. STS interactions occur when a large number of equivalent DNA sequences are simultaneously put together in solution, either before or after grafting on nanoparticles. The STS is more stringent than the SF constraint because it does not require a loop length for the hybridization of bases. This means that when STS is applied, SF is automatically fulfilled. Therefore, SF without STS interactions is relevant only in applications where unique strands are considered. In general, most experimental work is performed in liquid environments with an ensemble of interacting strands, for which STS interactions dominate. The third major constraint (PH), shown in Figure 2c, is to prevent the sticky-end sequence from partial hybridization when located in the presence of its complementary strand. This happens when similar short sequences are repeated several times within the same strand. Multiple PH is then possible with the complementary strand, along short portions of the ssST, weakening the binding of the two strands, compared to a total and more stable hybridization. For a proper implementation of the algorithm, a second input parameter needs to be provided by the user: the minimum number of adjacent complementary bases (noted as MNb) that must be avoided to prevent long-life SF and/or PH. Referring to Gibbs free

energies reported in the literature,<sup>38</sup> a default value of three bases has been assigned to this parameter. Other choices may be necessary for experiments using unconventional conditions, for example, at very low temperatures. Applications with MNb = 2 and MNb = 3 are discussed in the Methods section. Finally, the algorithm makes it possible to introduce MM constraints that can prevent folding on demand with more complex local sequences than just adjacent complementary bases. Panels d–f in Figure 2 show examples of forbidden sequences. Figure 2d shows a prohibited sequence of three adjacent complementary bases. Figure 2e,f shows examples of prohibited MM where one or two voids of noncomplementary bases are present within a hybridized local sequence. Figure 2e shows a sequence of four adjacent bases: two complementary, one noncomplementary, and one complementary. Figure 2f represents a sequence of six adjacent bases: two complementary, two noncomplementary, and two complementary.

Once all of the allowed sequences are built by respecting the constraints imposed by the user, the determination of their thermodynamic properties can then be used for a quantitative screening. In our calculations, the melting temperature is taken as the scoring parameter, but other physical properties can be easily substituted for the melting temperature for other purposes. The hybridization energies and melting temperatures are determined using the well-known Santa Lucia model with DNA and NaCl concentrations fixed at 1 mM and 1 M, respectively.<sup>38</sup> Note that performing the quantitative screening during the construction of the DNA strand is a simple method for reducing computing time. But one does not always know in advance the evolution of the quantitative physical properties as the DNA length is increased. Therefore, quantitative screening should not be performed on a set of sequences that is too restricted because this can discard interesting sequences from the beginning. Finally, we have not considered more complex constraints related to nonspecific interactions, such as wobble pairs (G-T) and G-quadruplexes,<sup>39,40</sup> although they could be simply implemented in our algorithm. The reason is double: (i) such interactions are, thermodynamically speaking, difficult to compare with well-established specific interactions and/or (ii) they do not substantially reduce the size of the subset of strands submitted to the final individual evaluation. In the particular case of G-quadruplexes, they appear only at strand lengths of around 20 bases where they can be individually evaluated during posttreatment, concomitantly with the melting temperatures.

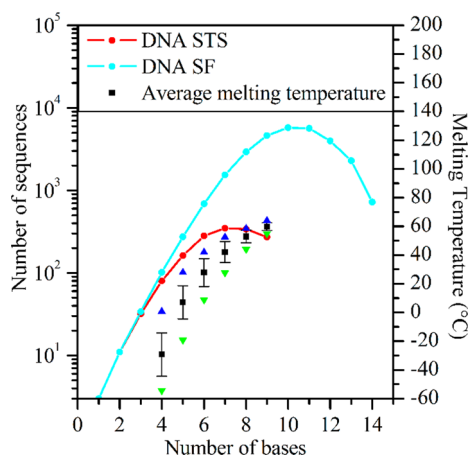
The algorithm is coded in C language and is available on demand, as is the complete list of sequences, for noncommercial purposes (<https://www.laas.fr/public/en/neo;rossi@laas.fr>).

## RESULTS AND DISCUSSION

**Illustration of the Calculation Method Applying Strong Constraints.** We first examine the simplest case where SF, STS, and PH constraints are considered with MNb = 2 and no MM. Even though this corresponds to poor interaction energies to stabilize associated folding configurations at room temperature, this very restrictive case gives us the opportunity to draw a clear step-by-step description of the method and to present details of the algorithm. In this simple case, the compatibility tensor is of rank two and is reduced to a  $4 \times 4$  matrix whose elements represent various sequences of two DNA bases made of the four A, C, G, and T nucleic acids. In this example, we assume a spacer made of a repeated sequence of thymine bases, which is best at enhancing the DNA grafting density as demonstrated before.<sup>15</sup> However, this condition is not mandatory and may be altered if needed in other realistic applications. (Other sequences with adenine spacers or without any DNA-based spacers are reported in Supporting Information SI-6 and SI-7.) Starting the construction after the thymine sequence, the matrix elements corresponding to TT and AA are set to 1 to forbid the

replication of any TT couples present and their complementary (AA) couple. Three choices are now possible for the first base—A, C, or G—because TT is forbidden. We consider each of them separately.

For example, if we choose A, then matrix elements TA and AT should also be set to 1, in addition to TT and AA. For the second base, only two choices—C and G—are possible because AT and AA are already forbidden. Another choice of the first base, C or G, does not necessarily lead to two possibilities for the second base. But globally, the number of possibilities does not systematically progress as  $4^N$ . On average, at each construction step, one or two matrix elements are set to 1. Therefore, one would expect that after  $n(n+1)/2$  steps, with  $n = 4$ , the whole matrix, representing the compatibility tensor, is set to 1, implying that there are no more allowed sequences. This behavior is represented in Figure 3 (red curve) with STS



**Figure 3.** Number of allowed sequences as a function of the number of bases in the sequence for  $MNb = 2$ , with STS and PH constraints (red curve) and folding and PH constraints (light-blue curve). Maximum (blue triangle), minimum (green upside down triangle), and average (black squares) melting temperatures calculated with DNA and salt concentrations set to 1 mM and 1 M, respectively. Standard deviations are also represented for average temperatures.

and PH constraints. Here, the number of allowed sequences is plotted against the length of the DNA strand. We observe that the number of sequences increases to 350 for strands that are 7 bases long and then decreases sharply down to zero for strands of 10 bases. The maximum of 350 sequences for 7-base DNA strands should be compared to 16 384 (or  $4^7$ ) possible sequences for DNA strands of the same length without any constraint. This reduction is drastically enhanced for more complex but realistic cases.

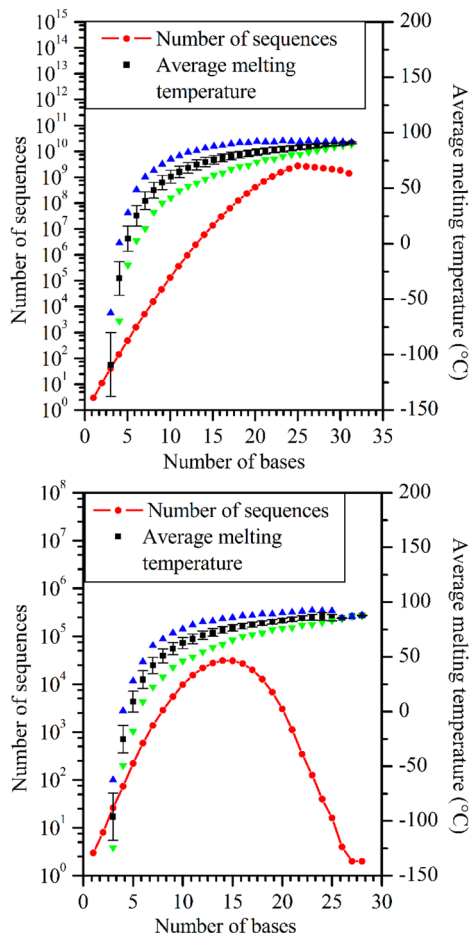
We conclude that, in this simple example, the maximum length of strands compatible with STS and PH constraints and  $MNb = 2$  is nine. The melting temperatures associated with the hybridized form of these short strands are very low, demonstrating reasonable stability at room temperature for sequences of only eight or nine bases. Relaxing the STS constraint and concentrating only on the SF and PH with  $MNb = 2$  leads to a drastic increase in the number of sequences and longer DNA strands as represented in Figure 3 (light-blue curve). We find a maximum of 5766 sequences for 10-base-long strands, out of more than  $10^6$  possible sequences, and a maximum length of 14 bases. However, STS is always present in most practical cases where an ensemble of strands is

considered in solution to react with a collection of nano-objects.

**Effect of Constraints on the Number of Bases and Melting Temperature.** A literature survey shows that most research teams use DNA strand lengths in the range of 4 to 40 bases without a consideration of DNA-base spacers. To reach these DNA strand lengths and consequently to increase the melting temperatures, we can increase the value of  $MNb$ , going beyond the simplest case described above. If we want to increase the length of the sequences, then we need to relax our initial set of constraints, which in turn will lead to increased folding or hybridization of intrinsic DNA in a noncontrolled manner. However, a slight relaxation of the presented constraints should not lead to technological difficulties because standard protocols are performed at temperatures at which nonpermanent hybridization is still tolerable (beyond possible exotic applications). To search for longer sequences beyond nine bases, we can relax the pairing limitation and adopt the value  $MNb = 3$ . To optimize an assembly protocol where multiple strands are used, we focus on constraints of interest in this application (i.e., STS and PH constraints).

Figure 4 shows the number of allowed sequences in two cases with  $MNb = 3$ , implying a compatibility tensor of rank three. In the top graph, STS and PH constraints are taken into account, and in the lower graph, selected MM constraints (as schematized in Figure 2e,f) are additionally included. Compared to the simplest case with  $MNb = 2$ , the number of allowed sequences increases drastically and longer sequences become also accessible. The reason is that it takes longer to populate tensors of rank three, containing 64 elements, than tensors of rank two, used in the previous and simpler example. For short sequences, we observe an exponential growth of the number of allowed sequences as a function of the number of bases. For the less restrictive case where MM constraints are not included, we find a maximum number of  $10^9$  allowed sequences for 25 base sequences out of  $10^{15}$  (or  $4^{25}$ ) possible sequences. The maximum length reaches 31 bases, which covers the main reported technological needs to date in terms of sequence length. Adding MM constraints drastically reduces the number of sequences: 31 018 allowed sequences for strands that are 14 bases long and a maximum length of 28 bases where only 2 sequences are allowed. We notice here that the restricted subset on which individual evaluation can be performed depends strongly on the imposed constraints. The actual subset to be considered is therefore a function of the complexity of the evaluation process. The largest set in the upper graph of Figure 4 is perfect for melting-temperature evaluation, which is a rapid calculation. However, it is largely oversized for thermodynamic equilibrium or kinetic evaluations, needing more computing resources.

After designing sequences corresponding to basic qualitative design rules, we turn to the determination of their thermodynamic properties for a quantitative screening, as stated in the Materials section. It should be noted that the number of allowed sequences is still very large. Further quantitative optimization is performed through a ranking procedure based on the melting temperatures. Other types of ranking procedures, based on physicochemical, thermodynamic, or kinetic properties, can be implemented in our software package but have not been used in the present investigation. Average as well as maximum and minimum melting temperatures are systematically calculated for all allowed sequence lengths and are reported in Figures 3 and 4. We notice that,



**Figure 4.** Number of allowed sequences as a function of the number of bases in the sequence for MNb = 3 (red curves). STS and PH constraints are taken into account in the top graph, and MM constraints are additionally included in the lower graph. Maximum (blue triangle), minimum (green upside down triangle), and average (black squares) melting temperatures calculated with DNA and salt concentrations set at 1 mM and 1 M respectively). Standard deviations are also represented for the average temperature.

independently of the constraints taken into account, the melting temperatures grow with an abrupt slope up to roughly 10 bases, where the average temperature is around 50 °C, before reaching a plateau. Temperatures just below 100 °C can then be obtained for sequences of 40 to 50 bases long. Interestingly, looking at the maximum melting temperatures, a saturation plateau is reached well before the average saturation value. The maximum melting temperatures saturates at around 92 °C for sequence lengths of roughly 20–25 bases. From a technological point of view, the maximum melting temperatures exhibit reasonable values starting with sequences constituted of 14 bases. Over 15 to 16 bases, there is no more notable evolution of the melting temperature because both the enthalpy and entropy terms simultaneously increase with the sequence length. Nevertheless, for technological purposes, increasing the sequence length may be beneficial to improving the mechanical properties of the DNA strands via their complete hybridization.

Thanks to the ranking algorithm, a list of optimized sequences can be obtained, as summarized in Table 1. The proposed list of sequences, for different strand lengths, is obtained with the most constrained optimization procedure

**Table 1. Best ssST Sequences, from 4 to 31 Bases, Resulting from a Ranking Procedure Based on the Melting Temperature<sup>a</sup>**

number of bases	sequence	melting temperature (°C)
4	GCGA	1
5	CCG CA	28
6	CCG CAC	42
7	ACC GCA G	52
8	ACG CAG GA	58
9 <sup>b</sup>	ACA GCC GAT	64
10	CCG TGG GCG A	75
11	AGC GGG TGC CT	79
12	AGC GGG TGC CTT	82
13	ACG GCA CCT CGC T	83
14	ACG GCA CCT CGC TT	85
15	AGC GGG TGC CTT GGA	86
16	ACC TTG GCT CCC GCA T	87
17	ACG ACA CCT TGC GGG CT	88
18	ACC CTG TGC TTC CGC CAA	89
19	ACC CTG TGC TTC CGC CAA C	89
20	AGG CGG TTG CTC TGG GAC AC	90
21	ACC CTG TGC TTC CGC CAA CGA	91
22	ACC CTG TGC TTC CGC CAA CGA G	92
23	ACC ACA GCG TCT CCT TGC CCG AA	92
24	ACC ACA GCG TCT CCT TGC CCG AAC	92
25	ACC ACA GCG TCT CCT TGC CCG AAC T	92
26	GAT AGG GAC TTC GTG TAA TGC TGG CG	85
27	CCT AAT GTA TCG CAG AGT TGA CGG GCT	87
28 <sup>c</sup>	GGT AAT CTA TGC GAC ACT TCA GCC CGT T	87
29	ACA ACG ACC ATC ACT CCC GCA GCC TTC TA	93
30	ACA ACC CGT CAC TAT GGA GCA GGC GAA GAT	92
31 <sup>d</sup>	ACA ACC ACG ACT CCC TAT GCC GCT GAA GAT T	91

<sup>a</sup>The proposed sequences take into account a spacer composed of seven repeated thymine bases. Additional sequences, including sequences preceded by either a neutral or an adenine spacer, is given in Supporting Information files SI-5 to SI-7. <sup>b</sup>From 4 to 9 bases, sequences emanate from calculations presented in Figure 3 (STS, PH). <sup>c</sup>From 10 to 28 bases, sequences emanate from calculations presented in Figure 4, lower graph (STS, PH, and mismatched). <sup>d</sup>From 29 to 31 bases, sequences emanate from calculations presented in Figure 4, top graph (STS, PH).

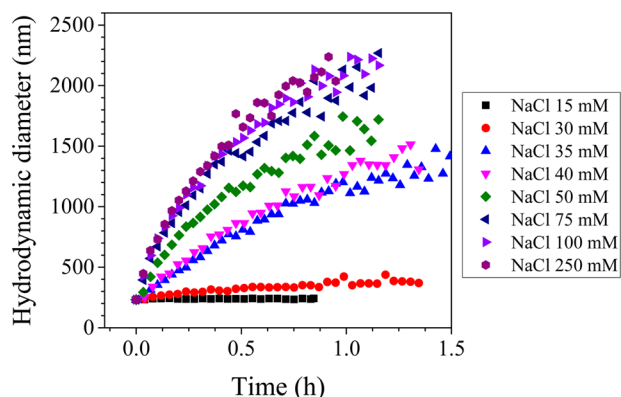
(i.e., both STS and PH constraints), as shown in the top panel of Figure 4. Note also that all sequences proposed for ssST in Table 1 take into account a spacer composed of seven repeated thymine bases.

In the Supporting Information file (Table SI-6), we derive optimized sequences with no spacer. (Note that the maximum length is 34.) These sequences would be more appropriate for technologies using pure chemical spacers, such as PEGs or alkyl chains. Note that using repeated adenine- instead of thymine-based spacers leads rigorously to equivalent sequence numbers: one has only to exchange A and T bases within the sequences. However, the melting temperature criteria may change the ranking procedure. Therefore, we also propose a list of best

candidates for the adenine spacer in the [Supporting Information](#) file (Table SI-7).

**Experimental Validation.** To experimentally test the relevance of our algorithm, we assembled Al and CuO NPs to obtain micrometer-sized aggregates of Al/CuO composites as described in the [Experimental Section](#) (method) to compare an optimized ssST (call ssA-Opt for Al nanoparticles and ssB-Opt for CuO nanoparticles) with a sequence of the same length proposed by Gang et al.<sup>21</sup> (called ssA-Lit for Al nanoparticles and ssB-Lit for CuO nanoparticles, respectively). A spacer composed of 15 repeated thymine bases is added for all DNA strand candidates for aggregation.<sup>21</sup>

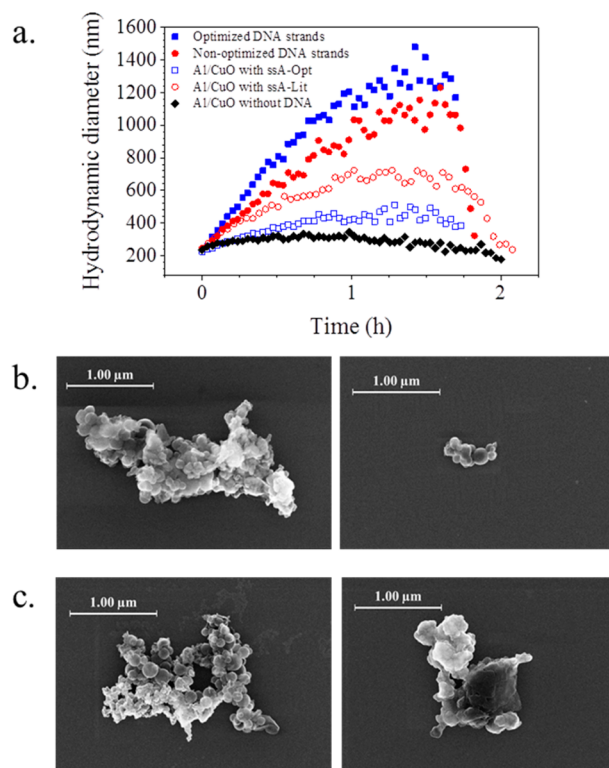
Because of negative repulsive charges of DNA grafted onto the nanoparticles, the addition of salt ( $\text{Na}^+\text{Cl}^-$ ) to the solution is necessary to screen these charges and allow the hybridization of ssST. Salt is therefore added to Al and CuO colloidal solutions at concentrations ranging from 15 to 250 mM while maintaining the same nanoparticle concentration. The aggregation kinetics are presented in [Figure 5](#). Two regimes



**Figure 5.** Evolution of the hydrodynamic diameter of colloidal suspensions of Al and CuO nanoparticles functionalized with ssA-Opt and ssB-Opt as a function of salt concentration, from 15 to 250 mM.

were monitored as a function of salt concentration, according to the literature.<sup>4,41,42</sup> A reaction-limited colloid aggregation (RLCA) exists from 30 to 50 mM in which the kinetics are slow and depend on the reaction between DNA strands. In this salt concentration range, the kinetics are highly dependent on the ionic strength of the solution. For salt concentrations above 75 mM, a diffusion-limited colloid aggregation (DLCA) regime is observed where the aggregation kinetics are no longer dependent on the salt concentration. In this regime, the ionic force dominates the repulsive force of nanoparticles, so the aggregation is driven by the probability of nanoparticle collision. From the kinetic study as a function of salt concentration, we chose a salt concentration of 35 mM because it exhibits the best profile in terms of full hybridization efficiency.

As seen in [Figure 6](#), the hydrodynamic diameter of Al and CuO nanoparticles before DNA functionalization remains almost constant, ranging from 250 and 300 nm (black curve), which confirms that Al or CuO nanoparticles do not aggregate under these conditions. SEM images related to this curve are given in [Supporting Information](#) SI-8. In a second experiment, CuO and Al nanoparticles were functionalized with a single noncomplementary sequence, with the optimized ssST taken from the literature. For both ssST's, aggregation is observed with a maximum diameter of 700 nm obtained for non-



**Figure 6.** (a) Evolution of the hydrodynamic diameter of colloidal suspensions of Al and CuO NPs as a function of time in PB with a NaCl concentration of 35 mM and Tween 20 at 0.05 vol % at pH 7 and 25 °C. The black curve corresponds to the hydrodynamic diameter of a mixture of Al and CuO before DNA functionalization. The curve with filled blue squares corresponds to a mixture of Al and CuO nanoparticles functionalized with ssA-Opt and ssB Opt, and the curve with empty blue squares corresponds to a mixture of Al and CuO both functionalized with ssA-Opt. The curve with filled red circles corresponds to a mixture of Al and CuO nanoparticles functionalized with ssA-Lit and ssB Lit, and the curve with empty red circles corresponds to a mixture of Al and CuO both functionalized with ssA-Lit. (b) SEM images of aggregates obtained with (left) complementary and optimized ssST and (right) noncomplementary and optimized ssST. (c) SEM images of aggregates obtained with (left) noncomplementary and nonoptimized ssST and (right) noncomplementary and optimized ssST.

optimized ssST, which is slightly higher than in the case of optimized ssST (~450 nm). In both noncomplementary regimes, the aggregation is driven by DNA-related nonspecific interactions and unwanted hybridizations as described later in the [Results and Discussion](#) section. But we see that the unwanted interactions are much more limited when DNA sequences are properly optimized by the algorithm, and that is why the aggregation is less pronounced for optimized strands.

In contrast, nanoparticles functionalized with complementary strands aggregate following a different regime and kinetics that depend on the ssST encoding sequences. Indeed, [Figure 6](#) (curve with filled blue squares) clearly shows a faster aggregation of nanoparticles functionalized with optimized ssST than nanoparticles functionalized with nonoptimized ssST (curve with filled red circles). After 1.5 h of incubation at room temperature, the average hydrodynamic diameter increases from 250 nm to 1.5  $\mu\text{m}$  when functionalized with optimized ssST whereas it increases from 250 nm to 1.2  $\mu\text{m}$  when the nanoparticles are functionalized with nonoptimized ssST. Interestingly, sedimentation is also observed for nonoptimized

ssST, despite a lower final hydrodynamic diameter compared to respective experiments with optimized ssST, under similar conditions and experimental time duration. This phenomenon is probably caused by a higher aggregate density compared to the optimized aggregate one. This is addressed in the [Results and Discussion](#) section.

Overall, DLS measurements are supported by the SEM images presented in [Figure 6b,c](#). Aggregates obtained with complementary and optimized ssST ([Figure 6b](#), left) are dense with a homogeneous composition of CuO and Al. The aggregates size is measured to be 1–3  $\mu\text{m}$ . However, aggregates obtained with complementary but nonoptimized ssST ([Figure 6c](#), left) have a different structure that is open and less dense. The arrangement of Al and CuO nanoparticles is less homogeneous.

Aggregates obtained with noncomplementary ssST ([Figure 6b,c](#), right) contain both Al and CuO nanoparticles with shapes very similar to those observed for Al and CuO nanopowders before the colloidal preparation and DNA functionalization (SI-1 and SI-2) [Figure 6b,c](#) (right images) clearly shows that some aggregation occurs depending on the type of ssST used. The aggregate size is evaluated to be  $\sim 800$  nm to 2  $\mu\text{m}$  when using nonoptimized ssST, whereas it is maximal at 500 nm when using optimized ssST.

**Discussion of the Experimental Results.** We begin the discussion with a closer examination of the sequence proposed by Gang et al. First, the sequence exhibits two types of anomalies compared to the optimized ssST. It includes one PH possibility with a non-negligible extent of four bases AGGT so that the complementary strand may hybridize incompletely (AAT-AGG-TGA-AGG-TTA). Second, when the spacer is composed of repeated thymines, an SF process of the sequence also becomes possible (...TT-AAT-AGG-TGA-AGG-TTA). Although this SF process may inhibit the hybridization interaction because of folding, it authorizes STS interactions when considering an ensemble of identical sequences. In conclusion, the nonoptimized sequence allows for multiple parasitic interaction possibilities whereas the optimized sequence does not (in the limit of two adjacent bases). In light of these characteristics, several conclusions can be drawn pointing to the importance of designing ssST properly. The most important is that, without complementary strands (empty squares and circles in [Figure 6](#)), aggregation saturates at a higher hydrodynamic diameter for the nonoptimized ssST than for the optimized ssST. In addition to the nonspecific DNA interactions that can be reasonably considered to be equivalent in both cases, the nonoptimized ssST offers the possibility of one STS interaction clearly causing interactions between nanoparticles of all types (as seen from SEM images). In the experiments with complementary strands (filled squares and circles in [Figure 6](#)), it is important to note that both curves include the effects seen in the noncomplementary experiment. For this reason, we can consider that almost all aggregation seen with the optimized ssST relies on full hybridization because no aggregation is observed in the noncomplementary experiment (only DNA nonspecific contributions). In contrast, with the nonoptimized ssST, we can determine the approximate amount of full and partial hybridization contributions among all other hybridization possibilities (strand folding, strand-to-strand hybridization) by subtracting the empty red circles from the filled red circles curves. Compared to the aggregation level of the optimized ssST (filled blue squares) consistent only with full hybridization, the non-

optimized ssST exhibits a dramatic lowering of hybridization (roughly 45 to 50% lowering) that leads to homogeneous Al/CuO mixing. Unfortunately, the presence of multiple hybridization contributions (full, PH, SF, and STS) in the nonoptimized ssST prevents a quantitative partitioning of relative individual impact of these hybridization defects on aggregate properties.

From a purely kinetic viewpoint, a quantification of the aggregation rate can be derived from DLS measurements and approximated by a linear law for the first hour. This gives, for the noncomplementary optimized and nonoptimized ssST (empty squares and circles in [Figure 6](#)),  $670 \pm 25$  and  $260 \pm 10$ , respectively. In contrast, the model experiment with complementary optimized and nonoptimized ssST (filled squares and circles in [Figure 6](#)) leads to much faster kinetics, the highest being in the case of the optimized ssST sequence (coefficients are  $1080 \pm 25$  and  $750 \pm 20$  for optimized ssST and nonoptimized ones, respectively).

From these thermodynamic and kinetic data and in the considered temperature and salt concentration, STS-related intermediates slow down the aggregation of the nonoptimized ssST. From a fundamental viewpoint, it is hard at this stage to understand and quantify the relation between hybridization type and aggregation kinetics. But more importantly, indirect proof of the hybridization optimization impact can still be made in the absence of clear evidence of major differences in nanoparticle morphologies. Another crucial aspect emanating from DLS measurements concerns the sedimentation of aggregates that are observed for nonoptimized ssST when optimized sequences are still in suspension. In general, we can attribute this phenomenon to the parasitic effect of STS interactions leading to denser aggregates. The fact that sedimentation is seen in the two sets of experiments, with complementary and noncomplementary ssST, indicates that STS causing CuO-CuO aggregation should dominate Al-Al aggregation because CuO particles are 2.3 times denser than Al particles. This is directly related to the DNA grafting density level on NP surfaces, their homogeneity, and aspects such as direct interactions of DNA with these surfaces. The level of aggregation observed with noncomplementary ssST favors CuO aggregation. Thus, it is reasonable to think that the full hybridization process going on in the complementary ssST experiment is a means to temper this effect, explaining the sedimentation at higher hydrodynamic diameters. These assumptions are not contradictory to SEM observations of noncomplementary experiments. Clearly, noncomplementary and nonoptimized ssST exhibits inhomogeneous aggregates ([Figure 6c](#), right). Rigorously, this inhomogeneity is present in the optimized ssST experiment, probably causing premature sedimentation. Nevertheless, it is hardly visible because of the large size of the aggregates.

To investigate in more detail these preliminary conclusions, we characterized the thermal stability of aggregates. Indeed, we showed ([Table 2](#)) that the sequence optimization implies an increase in the melting temperature, which means a higher stability of hybridized strands within a larger temperature range. After salt and DNA concentrations correction, the theoretical melting temperatures of optimized and nonoptimized sequences are respectively 64.2 and 46.1  $^{\circ}\text{C}$  under our experimental conditions (calculations from Santa Lucia's method<sup>38</sup>). To confirm this aspect of theoretical optimization, Al/CuO aggregates obtained with ss-Lit and ss-Opt underwent temperature cycles. After 30 min of aggregation at room temperature,

**Table 2. Sequences of Oligonucleotides Used in This Article<sup>a</sup>**

name	modification (5')	sequence (5' to 3')	theoretical melting temperature (°C)
ssA-Opt	biotin	(T) <sub>15</sub> -AGC-GGG-TGC-CTT-GGA	85.8
ssB-Opt	biotin	(T) <sub>15</sub> -TCC-AAG-GCA-CCC-GCT	
ssA-Lit	biotin	(T) <sub>15</sub> -AAT-AGG-TGA-AGG-TTA	69.8
ssB-Lit	biotin	(T) <sub>15</sub> -TAA-CCT-TCA-CCT-ATT	

<sup>a</sup>ssA and ssB refer to two complementary strands. The suffix Opt depicts DNA strands designed by the algorithm, and suffix Lit depicts DNA strands used in the literature.<sup>21</sup> The theoretical melting temperature is also calculated with DNA and salt concentrations set to 1 mM and 1 M, respectively.

each colloidal solution is heated to 50 or 70 °C (black and red curves, respectively) for 30 min and cooled to room temperature for 30 min. This cycle is repeated three times, and the hydrodynamic diameter is monitored and reported in Figure 7, where heating periods are represented by light-gray rectangles.

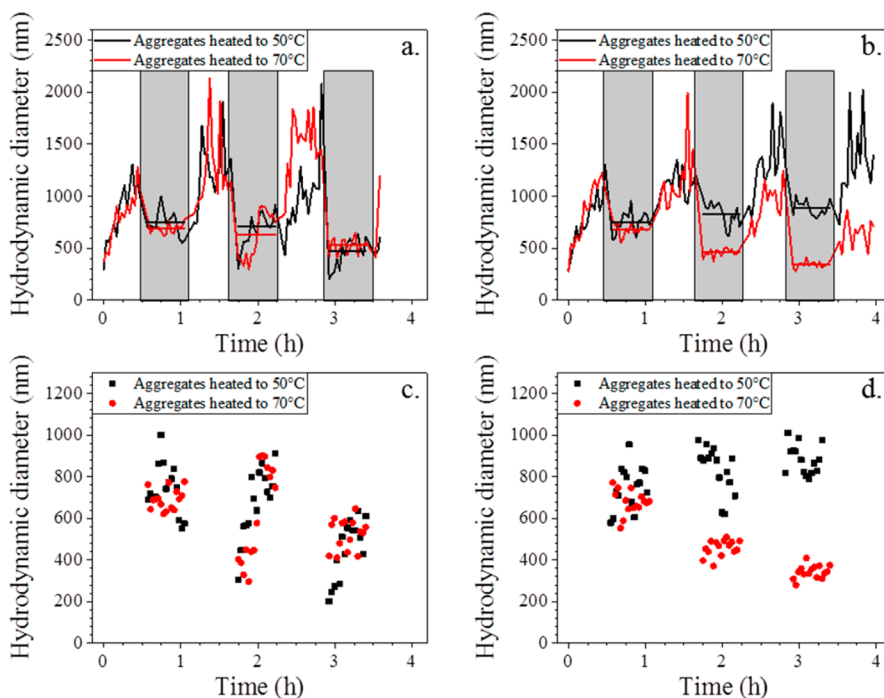
Figure 7a gives the evolution of the hydrodynamic diameter during heating at 50 and 70 °C for nonoptimized sequences. Figure 7b gives the evolution of the hydrodynamic diameter during heating at 50 and 70 °C for the optimized sequences. Figure 7c,d complements Figure 7a,b by plotting the hydrodynamic diameter data during only heating periods so that the data points can be plotted on an expanded y-axis scale.

For the first case (i.e., using the ssST that have been proposed in the literature), we observe a clear decrease in the hydrodynamic diameters after the first heating cycle for any

annealing temperature (straight lines in Figure 7a,b and data in SI-9). In contrast, when using optimized ssST, the integrity of aggregates is preserved at 50 °C (Figure 7b and black points in Figure 7d), whereas a deaggregation clearly appears at 70 °C (red points in Figure 7d). The baseline of heating phases for both red and black curves indicates that a temperature of 70 °C is necessary to efficiently disassemble the Al/CuO aggregate down to 300 nm while 50 °C heating only partially disassembles aggregates because the final size is ~900 nm. These experimental results clearly support the fact that the sequence melting temperature impacts the thermodynamics of aggregates.

However, note that the assembly kinetics are far more difficult to interpret by considering only DNA ssST because it is driven by multiple factors such as temperature, salt concentration, DNA density distribution, the distinct role of hybridization possibilities, and interactions between DNA and nanoparticles and between nanoparticles (colloidal stabilization). For instance, we demonstrated in previous work the high affinity of DNA bases with alumina surfaces.<sup>43</sup> Indeed, an examination of measurements during heating periods shown in Figure 7c,d clearly reveals a degradation of aggregation with heating cycles and also a contrast between optimized and nonoptimized ssST. After each heating cycle, the variation of aggregate hydrodynamic diameters between high temperature (50 and 70 °C) and room temperature is in a range of 40–60% because temperature may modify the DNA-based complexes at the surfaces. This variation is less pronounced for the optimized DNA ssST for which the amplitude between heating and cooling is more stable, indicating a better thermal stability and integrity of aggregates at both 50 and 70 °C.

Also, the equilibrated size dispersion measured over time for Al/CuO aggregates obtained with optimized DNA ssST



**Figure 7.** Evolution of the hydrodynamic diameters of Al/CuO aggregates functionalized with (a, c) nonoptimized complementary strands and (b, d) optimized complementary strands as a function of temperature. Aggregates were heated to 50 or 70 °C (black and red curves or points, respectively). Gray rectangles represent the period when the solution is heated. The other areas correspond to the period when the solution is cooled to room temperature. For all curves, the average hydrodynamic diameters are indicated at all heating periods by straight lines.

indicates that deaggregation follows a more homogeneous process. This is probably caused by the uniqueness of the process leading to more homogeneous and reproducible materials compared to the nonoptimized strand. This observation underscores the need to control the design of the ssST through our modeling procedure.

## CONCLUSIONS

In this article, we propose and describe an efficient and general algorithm that makes it possible to generate DNA single-strand sticky termination sequences with controlled folding and hybridization properties of interest in DNA nanotechnologies. We make use of this algorithm in the context of the DNA-directed assembly of nanoparticles. We analyze our simulation results as a function of constraints imposed on the design of DNA sequences, namely, self-folding and strand-to-strand interactions of the DNA strand and partial hybridization between complementary strands with additional mismatched constraints during hybridization. We show quantitatively how the combinatorial behavior of DNA with length can be overcome by our algorithm and measure the selective impact of these constraints. Using a ranking procedure based on the melting temperature determination, we provide sets of sequences from 4 to 34 base DNA ssST lengths to guide experimentalists in optimizing and better controlling their assembly protocols. The proposed sequences include common technological aspects such as the use of spacers (e.g., adenine or thymine or even non-DNA-based spacers) between the DNA sticky-end sequence and the surface, which have a direct impact on the optimization procedure. We then test this approach by developing a sequence that can be directly compared with the sequence found in the literature for DNA-directed Al and CuO nanoparticle assembly. The results demonstrate the beneficial impact of our methodology on the aggregation kinetics, thermal stability, and improved composition of these aggregates. Finally, we believe that our tool can motivate selective experimental studies relying on the generation of well-defined model ssST sequences with the objective of unraveling fundamental aspects of the relation between aggregation and DNA hybridization kinetics.

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: [10.1021/acs.langmuir.6b02843](https://doi.org/10.1021/acs.langmuir.6b02843).

Scanning electronic microscopy images, tables with sequences considered without spacers and with adenine spacers and the optimization of salt concentration (PDF)

## AUTHOR INFORMATION

### Corresponding Author

\*E-mail: [rossi@laas.fr](mailto:rossi@laas.fr).

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

We acknowledge financial support from the ANR/NSF CIREN (ANR/NSF-DMR-1312525), ANR IDEX Muse, the CNRS LIA-ATLAB initiative, the French Defense Agency (DGA), and Université Fédérale de Toulouse. We also acknowledge CALMIP computer resources.

## REFERENCES

- (1) Alivisatos, A. P.; Johnsson, K. P.; Peng, X.; Wilson, T. E.; Loweth, C. J.; Bruchez, M. P., Jr; Schultz, P. G. Organization of 'nanocrystal molecules' using DNA. *Nature* **1996**, *382*, 609–611.
- (2) Mirkin, C. A.; Letsinger, R. L.; Mucic, R. C.; Storhoff, J. J. A DNA-based method for rationally assembling nanoparticles into macroscopic materials. *Nature* **1996**, *382*, 607–609.
- (3) Brust, M.; Walker, M.; Bethell, D.; Schiffrin, D. J.; Whyman, R. Synthesis of thiol-derivatised gold nanoparticles in a two-phase Liquid–Liquid system. *J. Chem. Soc., Chem. Commun.* **1994**, 801–802.
- (4) Cobbe, S.; Connolly, S.; Ryan, D.; Nagle, L.; Eritja, R.; Fitzmaurice, D. DNA-Controlled Assembly of Protein-Modified Gold Nanocrystals. *J. Phys. Chem. B* **2003**, *107*, 470–477.
- (5) Castelino, K.; Kannan, B.; Majumdar, A. Characterization of Grafting Density and Binding Efficiency of DNA and Proteins on Gold Surfaces. *Langmuir* **2005**, *21*, 1956–1961.
- (6) Auyeung, E.; Morris, W.; Mondloch, J. E.; Hupp, J. T.; Farha, O. K.; Mirkin, C. A. Controlling Structure and Porosity in Catalytic Nanoparticle Superlattices with DNA. *J. Am. Chem. Soc.* **2015**, *137*, 1658–1662.
- (7) Daniel, M. C.; Astruc, D. Gold nanoparticles: assembly, supramolecular chemistry, quantum-size-related properties, and applications toward biology, catalysis, and nanotechnology. *Chem. Rev.* **2004**, *104*, 293–346.
- (8) Liu, J.; Lu, Y. A Colorimetric Lead Biosensor Using DNAzyme-Directed Assembly of Gold Nanoparticles. *J. Am. Chem. Soc.* **2003**, *125*, 6642–6643.
- (9) Ozin, G. A.; Arsenault, A. C.; Cademartiri, L. *Nanochemistry: A Chemical Approach to Nanomaterials*, 2nd ed.; Royal Society of Chemistry: Cambridge, U.K., 2009.
- (10) Liu, J.; Lu, Y. Fast Colorimetric Sensing of Adenosine and Cocaine Based on a General Sensor Design Involving Aptamers and Nanoparticles. *Angew. Chem.* **2006**, *118*, 96–100.
- (11) Park, D. J.; Zhang, C.; Ku, J. C.; Zhou, Y.; Schatz, G. C.; Mirkin, C. A. Plasmonic photonic crystals realized through DNA-programmable assembly. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 977–981.
- (12) Fan, J. A.; He, Y.; Bao, K.; Wu, C.; Bao, J.; Schade, N. B.; Manoharan, V. N.; Shvets, G.; Nordlander, P.; Liu, D. R.; Capasso, F. DNA-Enabled Self-Assembly of Plasmonic Nanoclusters. *Nano Lett.* **2011**, *11*, 4859–4864.
- (13) Séverac, F.; Alphonse, P.; Estève, A.; Bancaud, A.; Rossi, C. High-Energy Al/CuO Nanocomposites Obtained by DNA-Directed Assembly. *Adv. Funct. Mater.* **2012**, *22*, 323–329.
- (14) Herne, T. M.; Tarlov, M. J. Characterization of DNA Probes Immobilized on Gold Surfaces. *J. Am. Chem. Soc.* **1997**, *119*, 8916–8920.
- (15) Hurst, S. J.; Lytton-Jean, A. K. R.; Mirkin, C. A. Maximizing DNA Loading on a Range of Gold Nanoparticle Sizes. *Anal. Chem.* **2006**, *78*, 8313–8318.
- (16) Blythe, K. L.; Willets, K. A. Super-Resolution Imaging of Fluorophore-Labeled DNA Bound to Gold Nanoparticles: A Single-Molecule, Single-Particle Approach. *J. Phys. Chem. C* **2016**, *120*, 803–815.
- (17) Hill, H. D.; Macfarlane, R. J.; Senesi, A. J.; Lee, B.; Park, S. Y.; Mirkin, C. A. Controlling the Lattice Parameters of Gold Nanoparticle FCC Crystals with Duplex DNA Linkers. *Nano Lett.* **2008**, *8*, 2341–2344.
- (18) Park, S. Y.; Lytton-Jean, A. K.; Lee, B.; Weigand, S.; Schatz, G. C.; Mirkin, C. A. DNA-programmable nanoparticle crystallization. *Nature* **2008**, *451*, 553–556.
- (19) Maye, M. M.; Kumara, M. T.; Nykypanchuk, D.; Sherman, W. B.; Gang, O. Switching binary states of nanoparticle superlattices and dimer clusters by DNA strands. *Nat. Nanotechnol.* **2010**, *5*, 116–120.
- (20) Macfarlane, R. J.; Lee, B.; Jones, M. R.; Harris, N.; Schatz, G. C.; Mirkin, C. A. Nanoparticle Superlattice Engineering with DNA. *Science* **2011**, *334*, 204–208.
- (21) Zhang, Y.; Lu, F.; Yager, K. G.; van der Lelie, D.; Gang, O. A general strategy for the DNA-mediated self-assembly of functional

nanoparticles into heterogeneous systems. *Nat. Nanotechnol.* **2013**, *8*, 865–872.

(22) Auyeung, E.; Li, T. I.; Senesi, A. J.; Schmucker, A. L.; Pals, B. C.; de la Cruz, M. O.; Mirkin, C. A. DNA-mediated nanoparticle crystallization into Wulff polyhedra. *Nature* **2014**, *505*, 73–77.

(23) Jones, M. R.; Macfarlane, R. J.; Lee, B.; Zhang, J.; Young, K. L.; Senesi, A. J.; Mirkin, C. A. DNA-nanoparticle superlattices formed from anisotropic building blocks. *Nat. Mater.* **2010**, *9*, 913–917.

(24) O'Brien, M. N.; Jones, M. R.; Lee, B.; Mirkin, C. A. Anisotropic nanoparticle complementarity in DNA-mediated co-crystallization. *Nat. Mater.* **2015**, *14*, 833–839.

(25) Tkachenko, A. V. Morphological Diversity of DNA-Colloidal Self-Assembly. *Phys. Rev. Lett.* **2002**, *89*, 148303.

(26) Martinez-Veracoechea, F. J.; Mladek, B. M.; Tkachenko, A. V.; Frenkel, D. Design Rule for Colloidal Crystals of DNA-Functionalized Particles. *Phys. Rev. Lett.* **2011**, *107*, 45902.

(27) Padovan-Merhar, O.; Lara, F. V.; Starr, F. W. Stability of DNA-linked nanoparticle crystals: Effect of number of strands, core size, and rigidity of strand attachment. *J. Chem. Phys.* **2011**, *134*, 244701.

(28) Knorowski, C.; Burleigh, S.; Travesset, A. Dynamics and Statics of DNA-Programmable Nanoparticle Self-Assembly and Crystallization. *Phys. Rev. Lett.* **2011**, *106*, 215501.

(29) Knorowski, C.; Travesset, A. Self-Assembly and Crystallization of Hairy (*f*-Star) and DNA-Grafted Nanocubes. *J. Am. Chem. Soc.* **2014**, *136*, 653–659.

(30) Dhakal, S.; Kohlstedt, K. L.; Schatz, G. C.; Mirkin, C. A.; Olvera de la Cruz, M. Growth Dynamics for DNA-Guided Nanoparticle Crystallization. *ACS Nano* **2013**, *7*, 10948–10959.

(31) Li, T. I.; Sknepnek, R.; Olvera de la Cruz, M. Thermally Active Hybridization Drives the Crystallization of DNA-Functionalized Nanoparticles. *J. Am. Chem. Soc.* **2013**, *135*, 8535–8541.

(32) Vo, T.; Venkatasubramanian, V.; Kumar, S.; Srinivasan, B.; Pal, S.; Zhang, Y.; Gang, O. Stoichiometric control of DNA-grafted colloid self-assembly. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 4982–4987.

(33) Li, T. I.; Sknepnek, R.; Macfarlane, R. J.; Mirkin, C. A.; Olvera de la Cruz, M. Modeling the Crystallization of Spherical Nucleic Acid Nanoparticle Conjugates with Molecular Dynamics Simulations. *Nano Lett.* **2012**, *12*, 2509–2514.

(34) Seifpour, A.; Dahl, S. R.; Lin, B.; Jayaraman, A. Molecular simulation study of the assembly of DNA-functionalised nanoparticles: Effect of DNA strand sequence and composition. *Mol. Simul.* **2013**, *39*, 741–753.

(35) Pancoska, P.; Moravek, Z.; Moll, U. M. Rational design of DNA sequences for nanotechnology, microarrays and molecular computers using Eulerian graphs. *Nucleic Acids Res.* **2004**, *32*, 4630–4645.

(36) Wolfe, B. R.; Pierce, N. A. Sequence Design for a Test Tube of Interacting Nucleic Acid Strands. *ACS Synth. Biol.* **2015**, *4*, 1086–1100.

(37) Hagyard, T.; Williams, J. R. Potential of aluminium in aqueous chloride solutions. Part 1. *Trans. Faraday Soc.* **1961**, *57*, 2288–2294.

(38) SantaLucia, J.; Hicks, D. The Thermodynamics of DNA Structural Motifs. *Annu. Rev. Biophys. Biomol. Struct.* **2004**, *33*, 415–440.

(39) Mergny, J.-L.; Cian, A. D.; Ghelab, A.; Saccà, B.; Lacroix, L. Kinetics of tetramolecular quadruplexes. *Nucleic Acids Res.* **2005**, *33*, 81–94.

(40) Merkina, E. E.; Fox, K. R. Kinetic Stability of Intermolecular DNA Quadruplexes. *Biophys. J.* **2005**, *89*, 365–373.

(41) Lin, M. Y.; Lindsay, H. M.; Weitz, D. A.; Klein, R. C. B. R.; Ball, R. C.; Meakin, P. Universal diffusion-limited colloid aggregation. *J. Phys.: Condens. Matter* **1990**, *2*, 3093.

(42) Pamies, R.; Cifre, J. G. H.; Espín, V. F.; Collado-González, M.; Baños, F. G. D.; de la Torre, J. G. Aggregation behaviour of gold nanoparticles in saline aqueous media. *J. Nanopart. Res.* **2014**, *16*, 1–11.

(43) Calais, T.; Playe, B.; Ducéré, J. M.; Veyan, J. F.; Rupich, S.; Hemeryck, A.; Rouhani, M. D.; Rossi, C.; Chabal, Y. J.; Estève, A. Role of Alumina Coatings for Selective and Controlled Bonding of DNA on

Technologically Relevant Oxide Surfaces. *J. Phys. Chem. C* **2015**, *119*, 23527–23543.