Hierarchical multiscale simulation of DNA condensation

Tiedong Sun,[a] Alexander Mirzoev,[a] Nikolay Korolev,[a] Alexander P. Lyubartsev,[b,1] and Lars Nordenskiöld[a,1]

[a]School of Biological Sciences, Nanyang Technological University, Singapore 637551;

[b]Department of Materials and Environmental Chemistry, Stockholm University, 10691 Stockholm, Sweden

Author contributions: T.S. and A.M. performed simulations. T.S., A.M. and A.P.L. performed computational analyses. All authors designed the research, analysed the data, discussed the results wrote and approved the final version of the manuscript.

The authors declare no conflict of interest.

[1]To whom correspondence should be addressed: E-mail: larsnor@ntu.edu.sg (L.N.); alexander.lyubartsev@mmk.su.se (A.P.L.)

This article contains supporting information

DNA condensation at mesoscale level, induced by multivalent ions is of substantial importance for packing of DNA *in vivo* with many applications in biology, biotechnology and polymer physics. Rigorous modeling of this process with all-atom molecular dynamics (MD) simulations is presently impossible due to size and time scale limitations. Here, we present a hierarchical approach for systematic multiscale coarse-grained (CG) simulations of DNA condensation induced by the three-valent cobalt(III)-hexammine ($CoHex^{3+}$). On the basis of all-atom MD simulations, we extract solvent-mediated effective potentials for a CG model of DNA and simulate DNA aggregation in the presence of $CoHex^{3+}$. Further coarse-graining to a "super-CG" DNA model enables simulations of DNA condensation at mesoscale level. Modeling a 10 kbp-long DNA molecule results in formation of a toroid with distinct hexagonal packing in agreement with Cryo-EM observations. The approach uses no adjustable parameters and is applied on DNA up to megabase dimensions. It may be generalized to modeling chromatin up to chromosome size.

The compaction of DNA is a problem of outstanding importance in fundamental polymer and polyelectrolyte theory with many important applications in biology, biotechnology, nanoscience (1-4). While a long (~100 Mbp) chromosomal DNA molecule in low salt solution adopts a random coil conformation expanding over 100 μm, 46 such DNA molecules are packed inside the confined space of about 10 μm of the eukaryotic cell nucleus. Similarly, in sperm heads, bacteria and viruses, DNA is extremely densely condensed, packed into toroidal structures (5-8). DNA condensation has also attracted interest in gene delivery where the compaction is a key to optimizing such approaches.

For almost 50 years it has been known that *in vitro*, in the presence of highly charged cations like cobalt(III)-hexammine ($CoHex^{3+}$), $spermidine^{3+}$ and $spermine^{4+}$, DNA in solution condenses into collapsed structures of varying morphologies such as toroids, rod like fibres, globules and liquid crystals (9-13). While liquid crystalline phases are observed for 150 bp or shorter DNA molecules (14-16), long DNA molecules (a few to several hundred kbp) exhibit highly regular toroidal structures with DNA arranged in hexagonal packing inside the toroids (5), which have an outer diameter of around 100 nm, depending on conditions (17, 18). This spontaneous DNA toroid formation also observed *in vivo* in viruses and sperm chromatin, has fascinated scientists for a long time and been vastly studied experimentally with a variety of techniques such as X-ray diffraction (19, 20), Cryo-EM (7, 18) and more recently with single molecule techniques (21). This has resulted in significant advances in our understanding of the phenomenon both at mechanistic and fundamental level. However, there are still many unanswered problems related to DNA condensation in general and its spontaneous toroid formation induced by multivalent cations, in particular. Specifically, there is a lack of rigorous theoretical modeling approaches that are able to predict and reproduce this behavior from first principles.

Although counterintuitive, the fundamental origin of multivalent ion induced attraction between like charged DNA molecules leading to condensation is well established and grounded in the electrostatic properties of the highly charged DNA polyelectrolyte. Based on computer simulations as well as on analytical theories, it has been established that the attraction is caused mainly by ion-ion correlations which result in a correlated fluctuation in the instantaneous positions of the condensed counterions on DNA, which leads to a net attractive force between DNA molecules (reviewed in several works (2, 22, 23)). In the case of flexible multivalent cations like the polyamines or oligopeptides, the attraction is also generated by the "bridging" effect (24). However, although the origin of DNA-DNA attraction leading to condensation is clear and well described by different polyelectrolyte theories, DNA toroid formation is less well described theoretically. This phenomenon is important as an example of single molecule collapse of a semiflexible polyelectrolyte determined by a combination of DNA electrostatic interactions and mechanical bending properties. Commonly, computer modeling of DNA collapse and toroid formation (25-28) is based on a description of DNA as a chain of beads with parameterized harmonic bonds tuned to reproduce the DNA mechanical properties from experimental persistence length data. In such studies DNA-DNA attraction is modelled by effective empirical potentials and only in a few works electrostatic effects are treated explicitly by including ions in the simulation systems (23, 29-32). Common to most other approaches is that they treat generic polymer molecules and use empirical adjustable parameters to describe the relevant potentials in the models, where the connection to the atomistic DNA structure and chemical specificity is lost. While toroid-like structures are predicted, the hexagonal arrangement was only recently observed (28). In order to further advance our knowledge of this phenomenon and pave the ground for detailed analysis of the DNA condensation mechanism within various relevant

contexts such as compaction of DNA in chromatin, it is important to develop chemically specific DNA models without adjustable *ad-hoc* parameters.

The DNA toroid condensation is clearly intrinsic to the DNA molecule and inherent in its physico-chemical properties. The question arises whether this behavior can be predicted based on rigorous state-of-the-art all-atom molecular dynamics (MD) simulations. In recent years advances in computing technology have progressed considerably and all-atom biomolecular MD simulations including molecular water can now be performed for very large systems such as a nucleosome core particle and large DNA assemblies (33-35). However, simulation of DNA toroid condensation in all-atom detail over relevant time and length scales is not and will not be computationally feasible in a foreseeable future, and hence multiscale approaches linking atomistic and coarse-grained (CG) levels of description are necessary (36, 37).

Within a systematic bottom-up multiscale modeling scheme, the macromolecules are reduced to a CG description with effective sites, representing groups of atoms (38). The effective potentials between the CG sites are obtained from underlying atomistic simulations and thus implicitly include the effects of the specific atomic structure and water solvent (31, 39). The effective CG potentials can be computed using methods such as inverse Monte Carlo (IMC) (40), iterative Boltzmann inversion (41) or force matching (42), which allow to reduce the degrees of freedom number with several orders of magnitude, enabling simulations to be performed over a range of length and time scales. Although a few bottom-up (based on underlying MD simulations) approaches for multiscale modeling of DNA have recently emerged (reviewed in (36, 37)), they usually do not treat both electrostatics and solvent effects rigorously or do not reach the mesoscopic scale of DNA condensation.

Here, we perform systematic structure based multiscale CG simulations of DNA with explicit electrostatic interactions included, starting from all-atom description going up to

mesoscale modeling of DNA. We demonstrate DNA condensation induced by the three-valent $CoHex^{3+}$ ion forming toroids in hexagonal arrangement with a model obtained without adjustable parameters based only on the underlying all-atom force field and the topology adapted for the CG DNA description.

Results

Hierarchical multiscale approach. The stepwise multiscale hierarchical approach is illustrated in Fig. 1 and can be outlined as follows: First, using underlying Car-Parrinello MD (CPMD) optimized $CoHex^{3+}$ force field parameters (43), we perform all-atom MD simulations for a system consisting of four DNA oligonucleotides in the presence of $CoHex^{3+}$ (Fig 2). We then extract rigorous solvent mediated effective potentials by the IMC method (40) for a CG description of DNA, based on our previously validated model (44). The present CG model of DNA (Fig. 1B) is detailed in Fig. 1D (see Methods for more details).

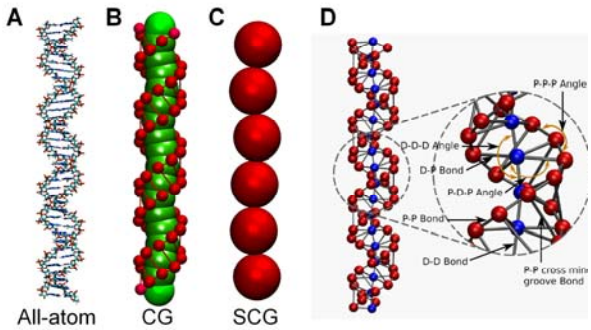

Fig. 1. DNA models at multiple scales. (**A**) The all-atom DNA model is the reference model. (**B**) The first level of coarse-graining provides the CG model, having two bead types. (**C**) The SCG model is a beads-on-string model built upon the CG model with one bead type. (**D**) The detailed presentation of the CG DNA model in (**B**). Four bond interactions, D-P, P-P (backbone), D-D and P-P (cross minor groove) are shown in grey. Three angle interactions (P-D-P, P-P-P and D-D-D) are indicated by orange arrows. The P-P cross minor groove bond is defined as the shortest P-P bond over the minor groove, which is P4(n)-P1(n+1), P3(n)-P2(n+1), where n is the unit number. For details, refer to the method section in the main text

Subsequently, we perform simulations with these potentials for a system comprising two hundred DNA molecules and explicit ions to simulate DNA aggregation in the presence of $CoHex^{3+}$. This simulation is used for further coarse-graining to a "super-CG" (SCG) DNA model (Fig. 1C) with another step of IMC. The derived effective potentials for the SCG model enable us to simulate DNA

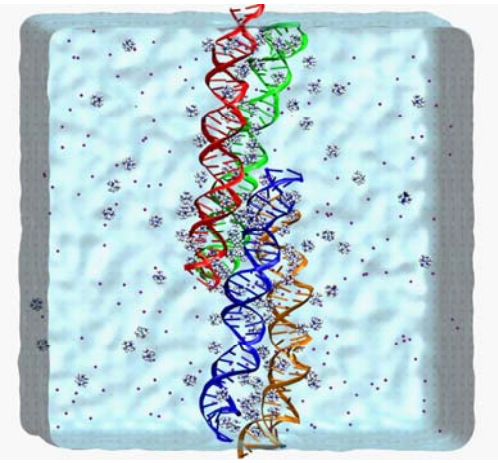

Fig. 2 The all-atom MD simulation was set up in a cubic box of 15 nm size with explicit water. The components of the box are: four 36 bp DNA double helices (coloured blue, red, green and orange, respectively), 140 $CoHex^{3+}$ ions (shown with liquorice representation), 140 K^+ ions, 95 Na^+ ions and 375 Cl^- ions. Water molecules are shown as blue surface; monovalent cations are drawn as purple dots. The snapshot in the figure represents an illustration at the end one of the three simulations, showing the four DNA molecules dandling together.

condensation at mesoscale level. We perform simulations for a system of hundreds of short DNA molecules (96 bp) as well as for a single DNA molecule (10 kbp).

DNA-DNA attraction in all-atom MD simulations. As a starting point for the bottom-up hierarchical approach modeling DNA condensation, we first performed all-atom MD simulations. The system with four DNA molecules in all-atom description containing explicit ions ($CoHex^{3+}$, Na^+, K^+ and Cl^-) and water is illustrated in Fig. 2. We made three independent 1 μs-long simulations that were used in the IMC procedure for extraction of effective solvent mediated potentials for the CG DNA model. Similarly to the result of our previous work (43), the system showed DNA-DNA attraction and aggregation of DNA into fibre-like bundles induced by the presence of $CoHex^{3+}$. This is illustrated in Fig. 2, which represents a snapshot at the end of one of the three trajectories. DNA fibres are formed across the periodic boundaries in all three independent simulations (see Supplementary Fig. S1). Snapshots at the end of the three final simulations show some variability in the character of DNA-DNA contacts over the periodic images, which may be due to the fact that even these microsecond simulations are not fully converged and bundling into various configurations may occur. Since CG models generally present a smoother free energy surface than its fine-grained MD counterpart, more efficient sampling can be achieved with CG models, which are also performed for much longer simulation times. A final snapshot from an

MD simulation of this four DNA system, but using the CG effective potentials (see below) is illustrated in the Supplementary Fig S1D. In the CG MD simulations a straight fibre configuration was easily reached and this result is stable and does not vary in different simulations.

Building a Coarse-Grained DNA model with effective solvent mediated potentials obtained from IMC. Next, following the all-atom MD simulations we proceed to extract the effective potentials for the CG model (illustrated in Fig. 1D) using the IMC procedure. The trajectories generated by the MD simulations were mapped from the all-atom to the CG representation and then used for calculation of RDFs and intramolecular (within DNA) distributions of bond lengths and angles between the CG sites, with averaging over all three independent trajectories. Supplementary Fig. S2 illustrates convergence in the IMC calculations during the iterations. Examples of calculated RDFs are shown in Fig. 3A-C. Following the IMC method, all effective interaction potentials for the CG DNA model were derived simultaneously within the same procedure, so that all correlations between different interaction modes are accounted for.

We additionally performed a control all-atom MD simulation without the presence of $CoHex^{3+}$ under conditions that should not lead to DNA condensation. The $CoHex^{3+}$ ions were replaced by the same number of Mg^{2+} ions (with the corresponding decrease of Cl^{-} counterions), resulting in the absence of DNA condensation, which is in agreement with the experimentally known fact that the divalent Mg(II) does not induce DNA condensation (45). We followed the same modeling protocol as for the case of $CoHex^{3+}$ to derive effective potentials for this control model. Selected distribution functions and final effective potentials are plotted in Figs. 3A-F. All RDFs and effective potentials for the CG simulations with $CoHex^{3+}$ are presented in the Supplementary Fig S3.

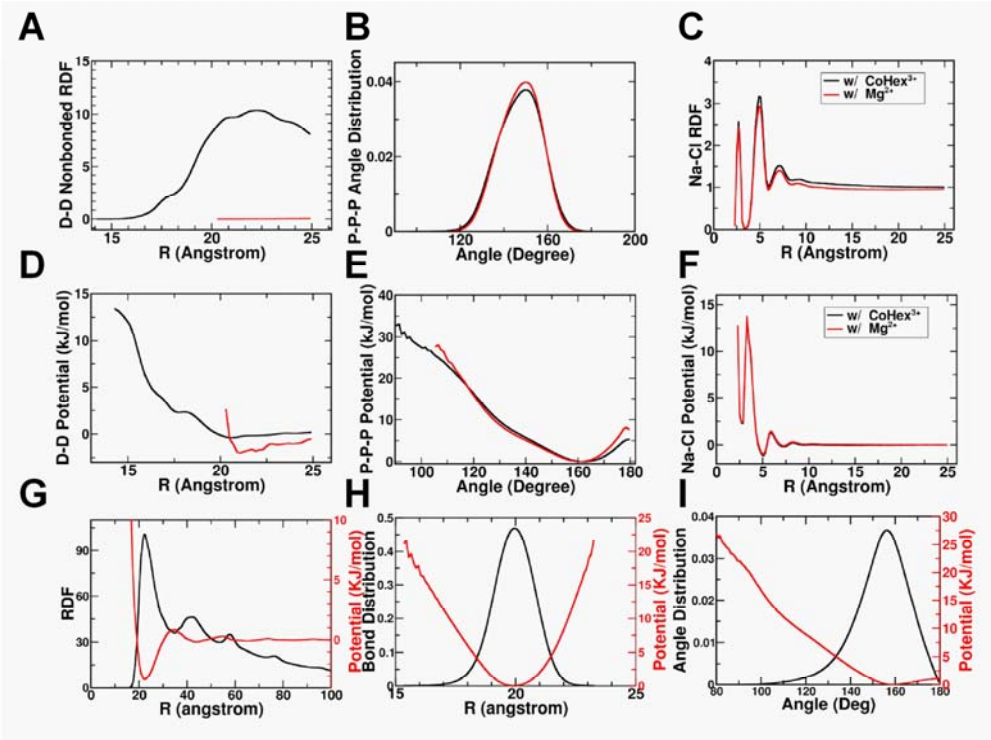

Fig. 3. Selected RDFs (top row), effective potentials of the CG model (middle row) and all RDFs and effective potentials of the SCG model (bottom row). In (**A-F**), three distributions and potentials are presented: D-D non-bonded pair, P-P-P angle along backbone and Na-Cl non-bonded pair, from both systems with $CoHex^{3+}$ (black) and Mg^{2+} (red). In (**G-I**), distributions and effective potentials of the SCG model are plotted together for the S-S non-bonded interaction (**G**), the S-S bond (**H**) and the S-S-S angle (**I**).

An important point from the plots in Fig. 3 should be noted. Even though there is a profound difference in the appearance of the non-bonded D-D RDF (between the central beads of different DNA molecules) comparing the DNA simulation in the presence of $CoHex^{3+}$ with the simulation with Mg^{2+}, it can be noted that the final effective potentials have qualitatively the same features. The D-D RDF from the simulation with $CoHex^{3+}$ (black line in Fig. 3A) has high values up to 10, reflecting close DNA contacts, while the D-D RDF from the simulation with Mg^{2+} stays well below 1 up to the cut-off distance. However, the final effective potentials for the D-D pair from both simulations have similar features. They are both slightly attractive from about 21 Å to the cut-off 25 Å. In the distance range below 21 Å, both effective potentials are repulsive. Furthermore, the effective potentials for ion-ion interactions are virtually the same as shown in Fig. 3C and 3F. In Fig. 3C there is a clear

difference in the Na-Cl RDF for the aggregating and non-aggregating simulations due to different average DNA-DNA distances. However, the final effective potentials from both systems are almost identical. We attribute this to the fact that correlations between the different interaction terms are well represented in the present systematic model. This behavior implies good transferability of the derived effective potentials.

DNA aggregation in the CG simulations. Having rigorously extracted effective potentials for the CG DNA model, we use them in large-scale simulations investigating DNA condensation induced by $CoHex^{3+}$. The applied CG approach treats long range electrostatic interactions explicitly with the presence of all mobile ions in the system. The total solvent-mediated interaction potential between all charged sites in the system is a sum of a Coulombic potential scaled by the dielectric permittivity of water (ε=78), and a short-range interaction (within the cut-off distance) determined within IMC procedure. This treatment of the long-range electrostatic interactions was validated in our previous work (46). Not only does this lead to a rigorous description of the important electrostatic interactions, but also enables the CG model to be used under varying ionic conditions.

After the significant reduction in the number of degrees of freedom by coarse-graining, we can easily simulate DNA condensation in box of size of 150×150×150 nm^3 for extended time. This is 1000 times larger volume compared to what is affordable for the all-atom MD simulations, which have used a box 15×15×15 nm^3. Here, 200 pieces of 100-bp CG DNA double helices are randomly placed in the box together with $CoHex^{3+}$, potassium and sodium ions as well as the appropriate amount of chloride ions.

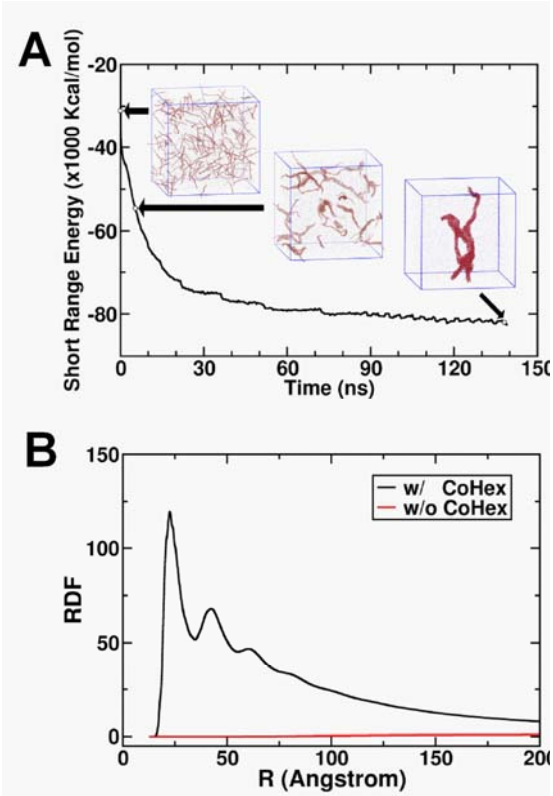

Fig. 4 (**A**) Snapshots and short range interaction energy profile from CG DNA simulation. Time points corresponding to each snapshot are indicated with black circles on the energy curve. DNA molecules are rendered red. (**B**) D-D RDF comparison between the CG simulations with and without $CoHex^{3+}$.

Figure 4A displays the short range energy of the CG model as a function of time with representative snapshots and illustrates that DNA aggregation occurs gradually during the simulation. Starting from a randomly dispersed distribution, DNA molecules gradually form contacts with each other. The DNA condensate particle gets larger, forming a single fibre-like particle at the end of the simulation where the DNA molecules demonstrate short-ranged hexagonal arrangement. The value of the smallest DNA-DNA distance is 22.5 Å, which is also exhibited by the first peak of the RDF in Fig. 4B. This is in reasonable agreement with experimental data although a longer separation may be expected for condensed short DNA. Our own observations (unpublished data) from X-ray diffraction measurements of precipitated 177 bp DNA molecules, induced by $CoHex^{3+}$, display a single broad Bragg peak at q = 0.26 $Å^{-1}$, with absence of long range hexagonal order. This corresponds to a lateral DNA-DNA separation in the range 24 Å – 27 Å (assuming hexagonal ($r = 4\pi/q\sqrt{3}$) or lamellar ($r = 2\pi/q$) packing). The shorter separation observed for bundled DNA in the simulations compared to experiment is likely due to the CHARMM force field used in the underlying all-atom MD simulations (47).

To test the robustness of the CG DNA model, we conduct another simulation where the $CoHex^{3+}$ ions are replaced by equivalent amount of K^+ and Na^+ ions using potentials obtained

from the IMC procedure described above. The same CG box size and simulation protocol are adopted. The system exhibits no DNA condensation and the DNA-DNA interaction is repulsive as can be seen from the D-D RDF plotted in Fig. 4B. In contrast to the system with $CoHex^{3+}$ amplitude of the D-D RDF is low in value, suggesting a large distance between DNA molecules. Hence, we can conclude that our CG DNA model is robust and produces realistic DNA aggregation behavior in large-scale simulations.

Building the SCG DNA model for mesoscopic-scale simulations. To investigate DNA condensation at the mesoscopic level we performed one more step of coarse-graining, constructing the super coarse-grain (SCG) model (Figs. 1C). The excellent behavior of the CG DNA model allowed us to confidently build a DNA model with even lower resolution and much better performance in terms of computational time. We performed the IMC using the trajectory obtained with CG DNA model. Three units (corresponding to 6 base pairs) of the CG DNA were mapped to one site (referred as "S") of this higher level SCG DNA model. After mapping, RDFs are extracted from the CG MD simulations. Since we only have one type of bead in the SCG model, the interactions consist of only one non-bonded term, one bonded S-S term and one S-S-S angular term. The ions were treated implicitly by the effective potentials, and the SCG DNA beads were uncharged, which reduced the computational costs of the model significantly. RDFs and effective potentials comprising the SCG DNA model are illustrated in Fig. 3G-I. Due to the simplicity of the model there is no significant correlation among the three interaction terms, the bond and angle potential minima are at the same position as the maxima in the corresponding respective distributions. These two terms may in principle be modelled by simple harmonic functions. On the other hand, the non-bonded interaction term cannot be directly fitted by conventional functions, such as a Lennard-Jones potential. Specifically, although there is a dominant minimum at 23

Å in the non-bonded effective potential, which might be mimicked by a Lennard-Jones potential, the long range behavior of the IMC-computed potential is different, with a positive maximum at 35 Å followed by two relatively small minima at about 44 Å and 65 Å. Hence, the final effective potential contains interaction features that preserve the characteristics of the underlying fine-grained CG simulation as well as the all-atom MD simulation. The present systematic hierarchical multiscale modeling approach can thus preserve more detailed information even with a DNA representation as simple as beads-on-a-string.

Mesoscopic simulations with the SCG model. Finally, the resulting SCG DNA model is used in mesoscale simulations of DNA condensation for two types of systems.

First, 400 relatively short DNA molecules, each one is equivalent to 96 bp, are randomly placed in a 150×150×150 nm^3 box. At the end of this simulation DNA condense into large particles consisting of more than 100 DNA molecules as illustrated in Fig. 5A. These particles exhibit a hexagonally ordered structure resembling a liquid crystal, illustrated in the cross-section view in Fig. 5B. In these particles, DNA molecules are arranged in such way that the hexagonal structure can be seen from the cross-section of the particle, as reported by experimental studies for short DNA molecules in the presence of the trivalent cations $CoHex^{3+}$ and $spermidine^{3+}$ (14, 15).

Secondly, we simulate a single 10 kb long DNA molecule, mimicking a dilute solution of long DNA molecules. The simulation starts from a fully extended DNA conformation. After a short relaxation at the beginning, a loop forms at one end of the DNA (see Supplementary Movie S1). The first snapshot in Fig. 5C shows the DNA conformation at this time point. Subsequently, the loop plays the role of a nucleation site, attracting more and more DNA beads to form a toroid that grows in size (see Supplementary Movie S2). Towards the end of the simulation, the whole DNA molecule is condensed into one toroidal particle (see

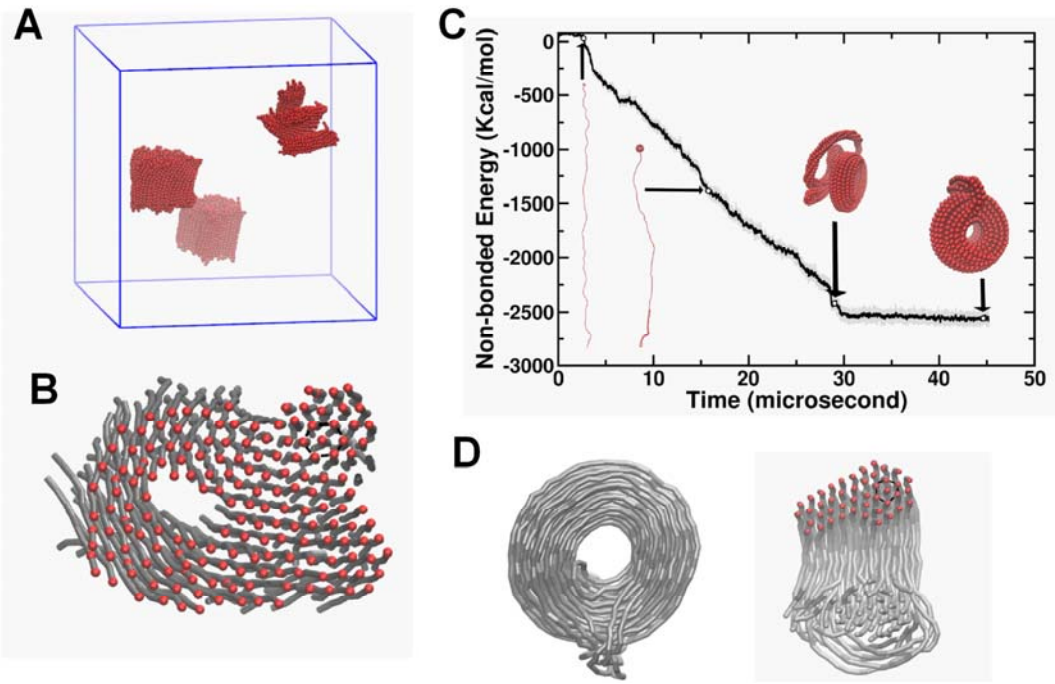

Fig. 5. DNA aggregation simulated by the SCG DNA model. (**A**) Final configuration of DNA aggregation in a simulation with 400 DNA molecules. (**B**) Cross-section of one of the DNA condensed particle shown in (**A**). (**C-D**) Formation of toroidal structures in 10 kbp DNA SCG simulations. (**C**) Energy profile and snapshots from one of the simulations. (**D**) Structure of the DNA toroid. Cartoon on the left hand side shows cross section through the toroid where the red dots illustrate DNA double helices near the cutting plane. A black hexagon is drawn to showcase the hexagonal arrangement of DNA in the toroid.

Supplementary Movie S3). The non-bonded interaction energy decreases as more and more DNA beads are involved in forming the toroid, illustrated by the energy profile in Fig.5C. Remarkably, as can be seen in Fig. 5D (right), the cross section of the toroid showed that DNA is organized in hexagonal arrangement. These structural features are consistent with the reported electron microscopy studies (5) and also predicted in a recent modeling study using empirical potential functions (28).

Analyzing multiple simulations reveals that toroids are mainly formed in two ways. The first scenario is by a single loop at one end initiating the toroid formation, while the other end may eventually form a fibre that subsequently joins the toroid at the end of the simulation (example in Supplementary movies S1, S2 and S3). Secondly, a loop may get formed at each end in the beginning of the simulation with the simultaneous growth of two toroids that eventually join together (exemplified in Supplementary movie S4). An interesting feature of

DNA toroid formation and size increase, is the sliding motion between contacting DNA segments. DNA toroids can adjust their conformation through this motion simultaneously with the growth due to rolling and attracting more DNA fragments. It should be pointed out that the shape of the toroids formed in the simulations may in several cases deviate significantly from an ideal toroid and appear cone-like. Occasionally, toroids are not formed but the final condensed structure is spheroid-like, which has also been reported in experimental EM observations (18). In the simulations, the toroid shape is, however, dominating. We also simulate lambda-DNA size single DNA molecules (48 kbp) that also form toroids, mainly by the mechanism with one nucleation loop at both ends (data not shown).

Next, we measure the dimensions of toroids formed by the 10kb DNA in multiple simulations shown in Fig. 6 (details are given in Supplementary Fig S4). The average thickness of the toroid is about 12 nm, which is smaller than in experiments for 3 kbp DNA (~25 nm) (18). Similarly, the observed diameter (~22 nm) and hole diameters (~10 nm) are also substantially smaller than the dimensions in the observed in the experiments (18). These values are reasonable given the differences between experiments and simulations. The major reason for smaller diameter and thickness in simulations as compared to experiment, is the fact that the simulations contain only a single DNA molecule. Experimentally the toroids formed from DNA of sizes below 50 kbp contain several DNA molecules which will lead to larger and thicker toroids (18). The diameter of the hole is

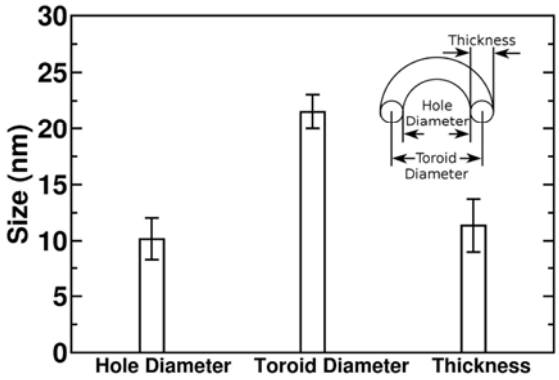

Fig. 6. Statistics of DNA toroid dimensions. Two parameters of the toroid are directly measured, hole diameter and outer diameter. The toroid diameter is taken as the arithmetic average of hole diameter and outer diameter. Toroid thickness is half the difference between hold diameter and outer diameter. Refer to Fig. S4 for details on measuring toroid dimensions.

expected to depend on nucleation loop size mechanism of toroid formation (18) and on the DNA bending properties as well as on effective DNA-DNA attraction. The observed small hole diameter is most certainly in part due to the intrinsic properties of the underlying force field, which compared to real DNA may represent a mechanically more flexible DNA with stronger attraction leading to shorter DNA-DNA distances and tighter packing in the toroid. The fact that toroid dimensions are strongly affected by electrostatic interactions between DNA molecules was earlier shown by Hud and co-workers (17), who demonstrated a pronounced dependence of toroid diameter and thickness as a function of ionic strength.

Discussion

In conclusion, we have developed a rigorous hierarchical multiscale simulation scheme, which enables simulation of DNA condensation at mesoscale levels. The phenomenon of multivalent induced DNA condensation is clearly inherent in the chemical properties of the DNA molecule. Inspired by this fact, we reasoned that a chemically based starting point, using state-of-the-art molecular force fields for all-atom MD simulations, followed by systematic coarse-graining, and using the IMC approach to extract solvent mediated effective CG potentials, would preserve those features of DNA in the CG models. Indeed, DNA condensation induced by the three-valent cobalt(III)-hexammine ions was demonstrated in large-scale simulations of hundreds of DNA molecules, which exhibited correct experimental structural features. We used a hierarchical approach where the CG model was further coarse-grained to a “super CG” model. Simulations at mesoscale level (10 kbp DNA) demonstrated toroid formation into hexagonally packed DNA, with reasonable dimensions in qualitative agreement with experimental observations. These results were obtained without any other underlying assumptions than the all-atom force field and the DNA topology model adopted in the CG simulations and without adjustable parameters.

In the present work we used all-atom MD simulations based on the CHARMM27 force field. However, we recently demonstrated similar behavior in all-atom simulations that showed DNA-DNA attraction and bundling using both CHARMM36 and AMBER parameters (43). This suggests that the results obtained in the CG simulations are also not force field dependent. Differences in e.g. DNA-DNA distances in the condensed system may occur, while we expect that the qualitative behavior would be predicted by both force fields. It should furthermore be of interest to investigate how the mesoscale simulation results depend on the CG topology comparing different CG DNA models, which include DNA sequence specificity (36, 48). It may also be noted that the present CG DNA model is not sequence specific, but such an extension can readily be implemented in the model (44).

The present successful approach lends support for developing CG models for more complicated systems exhibiting DNA compaction at mesoscale level such as chromatin and individual chromosomes. Such models can help predicting and understanding the compaction behavior of chromatin as a function of various variables known to regulate genome compaction such as histone tail modifications that change electrostatic interactions. Although multiscale modeling for nucleosomes and chromatin fibres following the present approach certainly increases the dimensionality of the CG system adding several degrees of freedom necessary to describe histones and their interactions, our present work along those lines shows that such extension is feasible (A. Mirzoev et al, unpublished).

Finally, in order to rigorously evaluate the time dynamics in the mesoscale simulations, generalized Langevin dynamics with friction parameters extracted from underlying detailed simulations can be performed, which enables the study of time-dependent condensation behavior (49, 50).

Methods

Systematic CG molecular simulations. Typically, the term ”systematic coarse-graining” in molecular modelling refers to building a low resolution molecular model based on properties presented by a high resolution model. The first steps of such a practice are usually empirical, which involve defining the system representation (so-called mapping rule) as well as choosing the functional form of Hamiltonian. This amounts to a process of reduction of the degrees of freedom. One should keep the most important degrees of freedom and only remove those of less importance but compromises would have to be made at this stage between an accurate representation and efficiency. The following step in coarse-grained modeling is to find the effective potentials. E.g., structure-based coarse-graining methods, such as inverse Monte Carlo (IMC) (40) and iterative Boltzmann inversion (IBI) (41) reproduce well the structural properties (space distribution functions etc.). Force-matching methods (42) can give accurate information on specific interactions, such as the potential of mean force.

All-atom Molecular Dynamics simulations. In the present work, we use atomistic force field model as our high resolution reference model and IMC as our method to extract effective potentials. Specifically, the coarse-graining is started with all-atom molecular dynamics simulations using the CHARMM27 force field (51, 52). In previous work (43) we have also tested the CHARMM36 and AMBER bsc0 which showed similar results concerning DNA aggregation in presence of $CoHex^{3+}$. We have settled on using the CHARMM27 force field in the present simulations, based on its better (compared to other force fields) agreement with experimental DNA persistence length behavior in CG DNA modeling (44). Here, the all-atom MD simulation is set up with four double helix DNA molecules, which are 36 bp long, placed in a cubic simulation box with periodic boundaries. The DNA sequence is the same as in our previous work (43) representing a 50-50% mixture of AT and CG pairs. The simulation box is large enough to avoid DNA self-interactions.

Cobalt(III)-hexammine ions, $CoHex^{3+}$, modelled as in our previous study are present to induce DNA condensation (43). The number of $CoHex^{3+}$ ions is determined in such way that the charge carried by $CoHex^{3+}$ is 1.5 times the charge of DNA, which should ensure attraction between DNA molecules. Additional salt is added to reach a salt concentration of 50 mM K^+ and 35 mM Na^+, with neutralizing amount of Cl- co-ions. The improved ion parameters by Yoo and Aksimentiev (53) are used throughout all simulations and the TIP3P water model was utilized. The system is illustrated in Fig. 2.

In total, three trajectories of 1 μs each are generated from the same starting configuration with DNA oligomers placed randomly in the simulation cell and different starting velocities. All bonds are constrained with the LINCS algorithm (54) implemented in GROMACS 5.1 (55), which enables a 2 fs time step. The system equilibration is conducted in three stages. First, DNA and $CoHex^{3+}$ ions are restrained while the system reaches a target temperature of 298 K and remains stable. This is followed by pressure coupling being turned on to maintain 1.013 bar pressure with only DNA molecules restrained, after which the unrestrained equilibration is conducted under constant temperature and constant pressure for 500 ns. Finally, the production phase is conducted for at least 500 ns. A velocity rescale thermostat (56) and Parrinello-Rahman barostat (57, 58) are adopted to regulate temperature and pressure respectively. Electrostatic interaction is treated with particle mesh Ewald (PME) (59) method with 10 Å real space cut-off. The van der Waals interaction is treated with a cut-off scheme with the potential shifted to zero at cut-off distance of 10 Å.

One control experiment is set up with the same all-atom simulation box, except that all $CoHex^{3+}$ ions are replaced by Mg^{2+} ions. Additional Na^+ and K^+ ions are added to keep charge neutrality. The simulation procedure is exactly the same as in the simulations with $CoHex^{3+}$.

Coarse-grained DNA models. We have performed coarse-graining at two spatial scales in order to reach a mesoscale level of DNA, resulting in two coarse-grained DNA models with different resolutions. The model at the first level is mapped from all-atom DNA (Fig. 1A), and is called the CG DNA model (Fig. 1B). The second model with lower resolution is called the super CG DNA (SCG) model since it is built on the first level CG DNA model (Fig. 1C). The CG DNA model has similar topology as our previous model, which was shown to well reproduce DNA persistence length dependence on salt concentration over a vast concentration range (44). Here, the double helical DNA is modelled with consecutive units of five beads, representing a two base pair fragment of DNA (illustrated in Fig. 1D). Among the five beads within each unit, four represent the phosphate groups (P), while the other one represents the four nucleosides in between (D). There are totally four types of bonds and three types of angles in the bonded interaction terms (Fig. 1D), which give rise to a helical structure with two distinguishable strands of phosphate groups where the major and minor grooves naturally appear from this topology (Fig. 1B). All ions are considered explicitly by assigning one CG site per ion.

The SCG DNA model is a beads-on-string model as shown in Fig. 1C. It consists of a string of beads, each representing three units (corresponding to six base pairs) in the CG DNA model. There is only one bead type (called ”S”) with zero charge. Bonded interaction terms are comprised of one bond type and one angle type. Compared to the CG DNA model, in the SCG model not only solvent is considered implicitly, but also ions, and electrostatic interactions are implicitly included into effective potentials between the S-beads.

Deriving effective potential by Inverse Monte Carlo. In the current CG DNA model, all effective interaction potentials are determined solely by the IMC method in a systematic and rigorous way. The only input information is the structural properties extracted from all-atom MD simulations in terms of the relevant RDFs between the sites of the CG model obtained by

mapping the all-atom MD trajectory from the four-DNA system to a corresponding CG site trajectory of the MD simulation. There are consequently no empirical parameter entering into this model and it rest solely on the all-atom DNA model and the corresponding CHARMM27 force field parameters as well on the CPMD optimized $CoHex^{3+}$ parameters.

To use IMC and derive the set of effective potentials defining the CG DNA model with explicit mobile ions, one radial distribution function (RDF) corresponding to each interaction term in the system is required. In order to avoid end-effect ”contamination” of the distribution functions, the ends of the DNA double helices in the IMC computations are treated as separate types (named DT and PT). Thus the total number of bead types in the system is eight: comprising four DNA beads (D, P, as well as terminal DT, PT), one of $CoHex^{3+}$, one of K^+, one Na^+ and one Cl^-; which gives the total number of non-bonded interaction terms equal to 36. We convert the all-atom MD trajectories from the three MD simulations into CG trajectories by applying the mapping rule described above. Convergence of DNA aggregation is confirmed by comparing DNA-DNA RDFs from trajectory segments at different simulation times. The final set of RDFs are calculated as averages over all three independent trajectories from the equilibrated sections of all three trajectories with equal weight (the first 0.5 μs were discarded). The total length of trajectories used is equal to 1.5 μs.

The IMC inversion calculation is carried out with the MagiC software (60) (which is also used for bead-mapping, RDF calculation, analysis and export of the resulting potentials). A zero potential is used as the first trial potential for non-bonded interactions, while the potential of mean force (defined as $U_{PMF} = -k_B T \ln(g(r))$) is used for bond and angle interactions. The effective potentials are refined in about 20 IMC iterations, with 100 parallel Monte Carlo sampling simulations in each iteration. In each sampling thread, 300 million Monte Carlo steps are performed with acceptance ratio maintained at about 50%. The first

half of each thread is considered as equilibration. The cut-off for RDF and non-electrostatic part of the effective CG potentials are set to 25 Å. Long range electrostatic interactions are treated using Ewald summation (59) with real space cut-off being 40 Å. The dielectric constant is set to 78.0. The Monte Carlo sampling is performed within a constant volume (equal to the average volume of the atomistic simulations) and constant temperature ensemble.

Coarse-grained MD simulations. The tabulated interaction potential for the CG DNA model, obtained as described above are then used in the CG MD simulations with a significantly bigger simulation box compared to the all-atom simulations. All MD simulations with the CG and SCG DNA models are conducted within the NVT ensemble. The LAMMPS (61) package is used for all CG/SCG MD simulations. For the CG simulations, the box of size is 150×150×150 nm^3. 200 pieces of 100-bp DNA double helices are randomly placed in such a simulation box together with $CoHex^{3+}$ ions. The total charge carried by the $CoHex^{3+}$ ions is twice the opposite charge on the DNA molecules. The simulation box also contains 10 mM potassium ions and 10 mM sodium ions as well as the appropriate amount of chloride ions. The CG simulation is started with a 1 fs time step to reach a stable temperature before switching to a 2 fs time step for next 1 million steps of equilibration. Langevin dynamics with damping parameter being 10 ps, is used to initiate the simulation. Finally, the production simulation is performed with 5 fs time step with a velocity rescale algorithm regulating system temperature. The particle-particle particle-mesh (PPPM) method is used to calculate electrostatic energies with a 40 Å real space cut-off. The same cut-off is applied for the short-range interactions.

The acquired trajectories are then used to map and build the SCG DNA model, following similar steps as building the CG DNA model and extracting effective potentials with IMC. In simulations with the SCG DNA model, the cut-off for short-range interactions is 200 Å. The

electrostatic interactions are not explicitly treated in the SCG model as they are effectively included into SCG potentials. In the simulations with multiple DNA, 400 pieces of 96 bp DNA (represented by a chain of 16 S-beads) are randomly placed in a 150×150×150 nm^3 box, serving as the starting configuration. The simulation is started with a randomly generated velocity at 298 K. Velocity rescale is used to regulate temperature in the first 10^5 steps to stabilize the temperature to 298 K. Finally the simulation is performed for 4×10^7 steps at the same temperature. The time step during the first 10^5 steps is 5 fs, while a 200 fs time step is used for the rest of the simulation. Furthermore, multiple simulations with single DNA molecule (consisting of 1700 beads, ~10kbp) are conducted to mimic dilute solution situation. In each simulation, one single DNA molecule is simulated in a 3450×3450×3450 nm^3 box, with no other components. The temperature of the first 10^5 steps of each simulation is kept at 298 K by velocity rescaling. After that, 2×10^8 steps are performed in each simulation at 298 K with a Langevin thermostat. The time step in the final production phase is 200 fs. The damping parameter of the Langevin thermostat is set to 100 ps to facilitate fast sampling.

Acknowledgments. We acknowledge the generous support of computer time allocation from the National Supercomputing Centre (NSCC) Singapore. We are indebted to Vishal Minhas for valuable discussions and to Prof Aatto Laaksonen for suggestions. This work was supported by the Singapore Ministry of Education Academic Research Fund (AcRF) Tier 2 (MOE2014-T2-1-123 (ARC51/14)) and Tier 3 (MOE2012-T3-1-001) grants (to L.N.) and by the Swedish Research Council (to A.P.L.).

1. Bloomfield VA (1996) DNA condensation. *Curr. Opin. Struct. Biol.* 6:334-341.
2. Gelbart WM, Bruinsma RF, Pincus PA, & Parsegian VA (2000) DNA-inspired electrostatics. *Physics Today* 53(9):38-44.
3. Carrivain P, *et al.* (2012) Electrostatics of DNA compaction in viruses, bacteria and eukaryotes: functional insights and evolutionary perspective. *Soft Matter* 8:9285-9301.

4. Thomas TJ & Thomas T (2018) Collapse of DNA in packaging and cellular transport. *Int. J. Biol. Macromol.* 109:36-48.
5. Hud NV & Downing KH (2001) Cryoelectron microscopy of lambda phage DNA condensates in vitreous ice: the fine structure of DNA toroids. *Proc. Natl. Acad. Sci. U.S.A.* 98(26):14925-14930.
6. Klimenko SM, Tikchonenko TI, & Andreev VM (1967) Packing of DNA in the head of bacteriophage T2. *J. Mol. Biol.* 23(3):523-IN521.
7. Leforestier A & Livolant F (2009) Structure of toroidal DNA collapsed inside the phage capsid. *Proc. Natl. Acad. Sci. U.S.A.* 106(23):9157-9162.
8. Ausió J, González-Romero R, & Woodcock CL (2014) Comparative structure of vertebrate sperm chromatin. *J. Struct. Biol.* 188(2):142-155.
9. Evdokimov YM, Platonov AL, Tikhonenko AS, & Varshavsky YM (1972) A compact form of double-stranded DNA in solution. *FEBS Lett.* 23(2):180-184.
10. Laemmli UK (1975) Characterization of DNA condensates induced by poly(ethylene oxide) and polylysine. *Proc. Natl. Acad. Sci. U.S.A.* 72(11):4288-4292.
11. Gosule LC & Schellman JA (1976) Compact form of DNA induced by spermidine. *Nature* 259:333.
12. Chattoraj DK, Gosule LC, & Schellman JA (1978) DNA condensation with polyamines: II. Electron microscopic studies. *J. Mol. Biol.* 121(3):327-337.
13. Wilson RW & Bloomfield VA (1979) Counterion-induced condensation of deoxyribonucleic acid. A light-scattering study. *Biochemistry* 18(11):2192-2196.
14. Sikorav JL, Pelta J, & Livolant F (1994) A liquid crystalline phase in spermidine-condensed DNA. *Biophys. J.* 67(4):1387-1392.
15. Pelta J, Livolant F, & Sikorav JL (1996) DNA aggregation induced by polyamines and cobalthexamine. *J. Biol. Chem.* 271(10):5656-5662.
16. Nakata M, *et al.* (2007) End-to-end stacking and liquid crystal condensation of 6 to 20 base pair DNA duplexes. *Science* 318(5854):1276-1279.
17. Conwell CC, Vilfan ID, & Hud NV (2003) Controlling the size of nanoscale toroidal DNA condensates with static curvature and ionic strength. *Proc. Natl. Acad. Sci. U.S.A.* 100(16):9296-9301.
18. Hud NV & Vilfan ID (2005) Toroidal DNA condensates: Unraveling the fine structure and the role of nucleation in determining size. *Annu. Rev. Biophys. Biomol. Struct.* 34:295-318.
19. Schellman JA & Parthasarathy N (1984) X-ray diffraction studies on cation-collapsed DNA. *J. Mol. Biol.* 175:313-329.
20. Rau DC & Parsegian VA (1992) Direct measurement of the intermolecular forces between counterion-condensed DNA double helices. Evidence for long-range attractive forces. *Biophys. J.* 61:246-259.
21. Brewer LR (2011) Deciphering the structure of DNA toroids. *Integr. Biol. (Camb).* 3(5):540-547.
22. Nordenskiöld L, Lyubartsev AP, & Korolev N (2008) DNA-DNA interaction. *DNA Interactions with Polymers and Surfactants*, eds Dias RS & Lindman B (John Wiley & Sons, Inc., London), pp 209-237.
23. Korolev N, Lyubartsev AP, & Nordenskiöld L (2010) Cation-induced polyelectrolyte-polyelectrolyte attraction in solutions of DNA and nucleosome core particles. *Adv. Colloid Interface Sci.* 158:32–47.
24. Podgornik R & Licer M (2006) Polyelectrolyte bridging interactions between charged macromolecules. *Curr. Opin. Colloid Interface Sci.* 11:273–279.
25. Yoshikawa Y, Yoshikawa K, & Kanbe T (1999) Formation of a giant toroid from long duplex DNA. *Langmuir* 15(12):4085-4088.
26. Stevens MJ (2001) Simple simulations of DNA condensation. *Biophys. J.* 80:130-139.
27. Ou Z & Muthukumar M (2005) Langevin dynamics of semiflexible polyelectrolytes: Rod-toroid–globule-coil structures and counterion distribution. *J. Chem. Phys.* 123(7):074905.
28. Dey A & Reddy G (2017) Toroidal condensates by semiflexible polymer chains: Insights into nucleation, growth and packing defects. *J. Phys. Chem. B* 121(39):9291-9301.
29. Lyubartsev AP & Nordenskiöld L (1997) Monte Carlo simulation study of DNA polyelectrolyte properties in the presence of multivalent polyamine ions. *J. Phys. Chem. B* 101(21):4335-4342.
30. Fan Y, Korolev N, Lyubartsev AP, & Nordenskiöld L (2013) An advanced coarse-grained nucleosome core particle model for computer simulations of nucleosome-nucleosome interactions under varying ionic conditions. *PLoS One* 8(2):e54228.
31. Hinckley DM & de Pablo JJ (2015) Coarse-Grained ions for nucleic acid modeling. *J. Chem. Theory Comput.* 11(11):5436-5446.
32. Córdoba A, Hinckley DM, Lequieu J, & de Pablo JJ (2017) A molecular view of the dynamics of dsDNA packing inside viral capsids in the presence of ions. *Biophys. J.* 112(7):1302-1315.
33. Yoo J & Aksimentiev A (2016) The structure and intermolecular forces of DNA condensates. *Nucleic Acids Res.* 44(5):2036-2046.
34. Ishida H & Kono H (2017) H4 tails potentially produce the diversity in the orientation of two nucleosomes. *Biophys. J.* 113:978–990.

35. Saurabh S, Glaser MA, Lansac Y, & Maiti PK (2016) Atomistic simulation of stacked nucleosome core particles: Tail bridging, the H4 tail and effect of hydrophobic forces. *J. Phys. Chem. B* 120(12):3048–3060.
36. Dans PD, Walther J, Gomez H, & Orozco M (2016) Multiscale simulation of DNA. *Curr. Opin. Struct. Biol.* 37:29–45.
37. de Pablo JJ (2011) Coarse-grained simulations of macromolecules: from DNA to nanocomposites. *Annu. Rev. Phys. Chem.* 62:555-574.
38. Takada S (2012) Coarse-grained molecular simulations of large biomolecules. *Curr. Opin. Struct. Biol.* 22(2):130-137.
39. Lyubartsev AP & Laaksonen A (2004) On the reduction of molecular degrees of freedom in computer simulations. *Lect. Notes Phys.* 640:219-244.
40. Lyubartsev AP & Laaksonen A (1995) Calculation of effective interaction potentials from radial distribution functions: A reverse Monte Carlo approach. *Phys. Rev. E* 52: 3730–3737.
41. Reith D, Putz M, & Muller-Plathe F (2003) Derived effective mesoscale potentials from atomistic simulations. *J. Comp. Chem.* 24:1624-1636.
42. Izvekov S & Voth GA (2005) Multiscale coarse-graining method for biomolecular systems. *J. Phys. Chem. B* 109:2469–2473.
43. Sun T, Mirzoev A, Korolev N, Lyubartsev AP, & Nordenskiöld L (2017) All-atom MD simulation of DNA condensation using ab initio derived force field parameters of cobalt(III)-hexammine. *J. Phys. Chem. B* 121(33):7761-7770.
44. Korolev N, Di L, Lyubartsev AP, & Nordenskiöld L (2014) A coarse-grained DNA model parameterized from atomistic simulations by inverse Monte Carlo. *Polymers* 6(6):1655-1675.
45. Bloomfield VA (1997) DNA condensation by multivalent cations. *Biopolymers* 44:269-282.
46. Mirzoev AA & Lyubartsev AP (2011) Effective solvent-mediated interaction potentials of Na+ and Cl- in aqueous solution: temperature dependence. *Phys. Chem. Chem. Phys.* 13:5722-5727.
47. Yoo J & Aksimentiev A (2018) New tricks for old dogs: improving the accuracy of biomolecular force fields by pair-specific corrections to non-bonded interactions. *Phys. Chem. Chem. Phys.* 20(13):8432-8449.
48. Hinckley DM, Freeman GS, Whitmer JK, & de Pablo JJ (2013) An experimentally-informed coarse-grained 3-site-per-nucleotide model of DNA: Structure, thermodynamics, and dynamics of hybridization. *J. Chem. Phys.* 139(14):144903.
49. Hijón C, Español P, Vanden-Eijnden E, & Delgado-Buscalioni R (2010) Mori-Zwanzig formalism as a practical computational tool. *Faraday Discuss.* 144:301-322.
50. Lei H, Caswell B, & Karniadakis GE (2010) Direct construction of mesoscopic models from microscopic simulations. *Phys. Rev. E* 81(2):026704.
51. Foloppe N & MacKerell AD, Jr. (2000) All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *J. Comp. Chem.* 21:86-104.
52. MacKerell AD, Jr. & Banavali N (2000) All-atom empirical force field for nucleic acids: II. Application to molecular dynamics simulations of DNA and RNA in solution. *J. Comp. Chem.* 21:105-120.
53. Yoo J & Aksimentiev A (2012) Improved parametrization of Li^+, Na^+, K^+, and Mg^{2+} ions for all-atom Molecular Dynamics simulations of nucleic acid systems. *J. Phys. Chem. Lett.* 3:45-50.
54. Hess B (2008) P-LINCS: A parallel linear constraint solver for Molecular Simulation. *J. Chem. Theory Comput.* 4(1):116-122.
55. Abraham MJ, *et al.* (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1–2:19-25.
56. Bussi G, Donadio D, & Parrinello M (2007) Canonical sampling through velocity rescaling. *J. Chem. Phys.* 126(1):014101.
57. Parrinello M & Rahman A (1980) Crystal structure and pair potentials: A Molecular-Dynamics study. *Phys. Rev. Lett.* 45(14):1196-1199.
58. Parrinello M & Rahman A (1981) Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* 52(12):7182-7190.
59. Essmann U, *et al.* (1995) A smooth particle mesh Ewald method. *J. Chem. Phys.* 103:8577-8593.
60. Mirzoev A & Lyubartsev AP (2013) MagiC: Software package for multiscale modeling. *J. Chem. Theory Comput.* 9:1512-1520.
61. Plimpton S (1995) Fast parallel algorithms for short-range Molecular Dynamics. *J. Comp. Phys.* 117:1-19.

Supplementary Materials

Hierarchical multiscale simulation of DNA condensation

Tiedong Sun, Alexander Mirzoev, Nikolay Korolev, Alexander Lyubartsev, Lars Nordenskiöld

1. Convergence of the Inverse Monte Carlo

Fig. S1 demonstrates convergence of the IMC procedure. We show examples of one intermolecular interaction (between $CoHex^{3+}$ and K^{+} ions) and one intramolecular interaction (between DNA phosphates across the minor groove), plotting the corresponding distribution functions and interaction potentials obtained at several iterations of the IMC procedure.

.

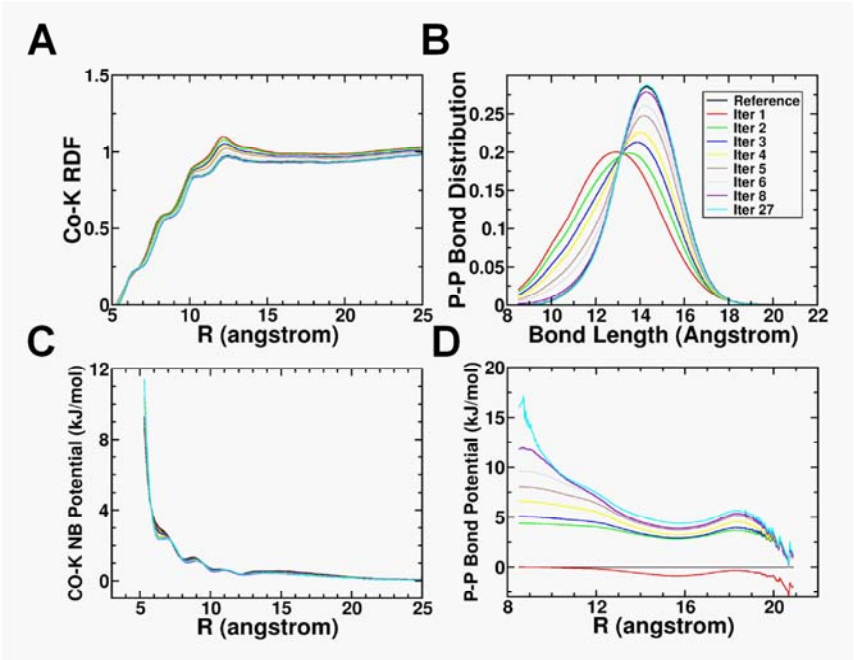

Fig. S1. Selected RDF (**A, B**) and effective potential (**C, D**) plots from several iterations in the IMC calculation. RDFs and potentials of one non-bonded term (Co-K, plots in **A** and **C**) and one bonded term (P-P cross minor groove bond, plots in **B** and **D**) are selected to illustrate the convergence of the RDF and the effective potential. The RDF underwent significant adjustments in the first few iterations of the runs. The adjustment in the potential progressively becomes smaller and smaller. The final potential, reproducing the whole set of reference distribution functions within the computational uncertainty, is achieved after 27 iterations.

2. The complete set of RDF and final effective potentials defining the CG model

Here we display RDFs and effective potentials corresponding to all interaction terms in the CG-model simulations with $CoHex^{3+}$. We have 4 bead types in DNA (DT, PT, D, P), as well as 4 types of ions (CO, K, NA, CL). The complete set consists of 36 non-bonded interaction terms, 4 bonded interaction terms and 3 angle interaction terms. By examining these plots, we can see the effect of correlations among interaction terms. As an example, there would be no potential minimum for the D-P non-bonded interaction at 6.5 Å if not for correlations from other interactions, since there is no peak at 6.5 Å in the D-P RDF.

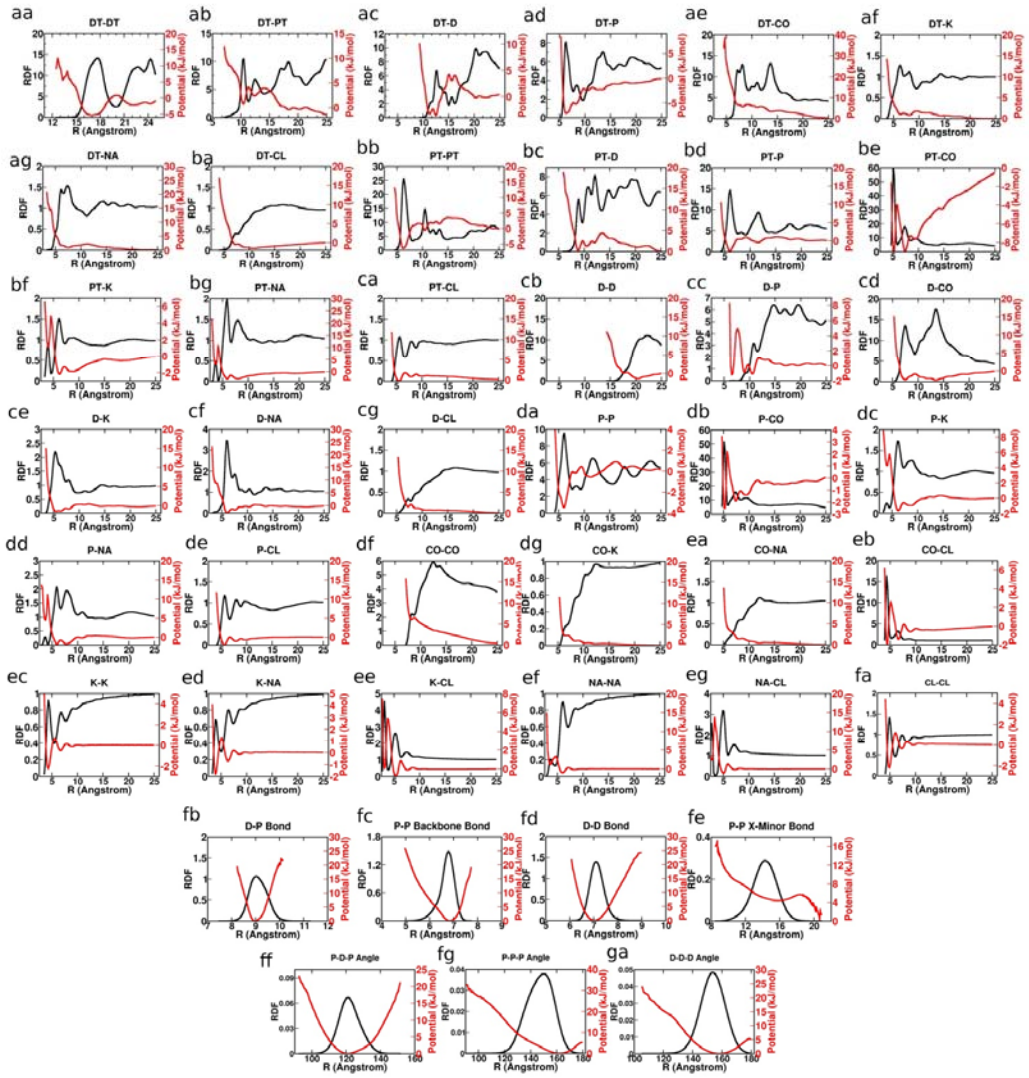

Fig. S2. RDFs (black) and the final effective potentials (red) for all interaction terms in the CG simulations with $CoHex^{3+}$. A total of 36 non-bonded interactions (**aa – fa**), 4 bond interactions (**fb – fe**) and 3 angle interactions (**ff – ga**) are shown.

3. Example of snapshots from different all-atom MD simulations

Because of the higher number of degrees of freedom in all-atom simulations, its free energy surface is more complicated than that of CG model. Due to the finite sampling time, the simulation may sample local minimum and display variable configurations of the bundled DNA molecules that interact over the periodic boundaries. Figure S3 (A-C) illustrates variability in final conformations from the three all-atom simulations.

On the other hand, in CG simulations, the free energy surface is much simpler and simulations can be significantly longer. A stable low energy state is easily reached and shows no variability in different simulations (Fig. S3D).

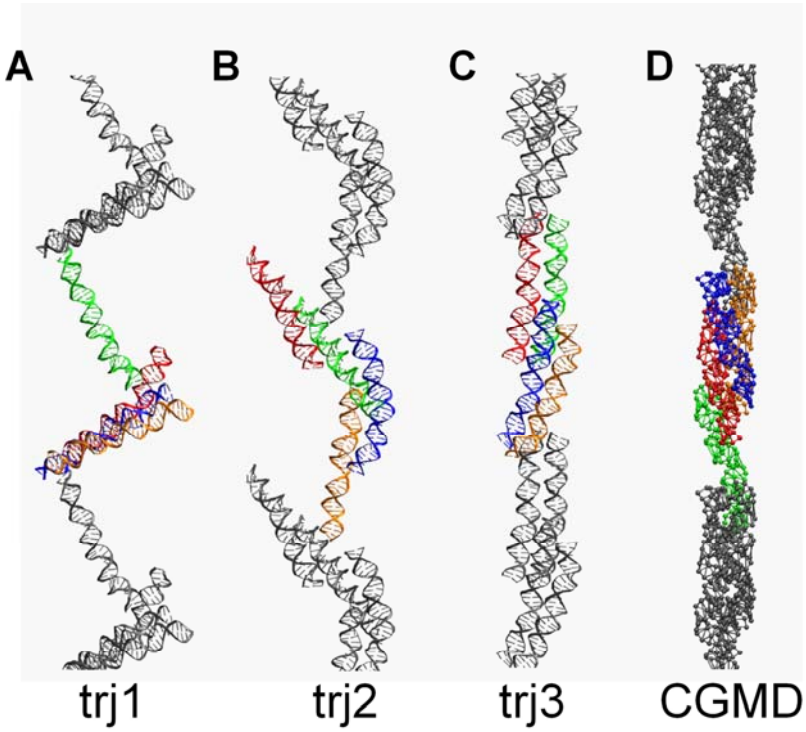

Fig. S3. Final conformations of all-atom trajectories (**A-C**). DNA double helices are colored differently. Periodic images of DNA double helices are colored grey. (**D**) Illustrates the final conformation in an MD simulation for the same system as in (**A-C**) but with the CG potentials.

4. Measurement of toroid dimensions

To determine the size of toroids that are formed in the simulations with the SCG model, multiple single DNA simulations have been conducted. The resulting toroids are measured by the following method, which tries to mimic the toroid measurement practice in EM experiments. First, the toroid is projected to a plane along the axes parallel to its hole. The projected density map is used to determine two toroid dimension parameters, hole diameter and outer diameter. As shown in Fig. S4B, two circles are used to represent the toroid hole and the outer dimension. The diameters of these two circles are recorded. Finally, the toroid diameter is determined to be the arithmetic average of the diameters of the two circles. Toroid thickness is defined as half the difference of the diameters of two circles.

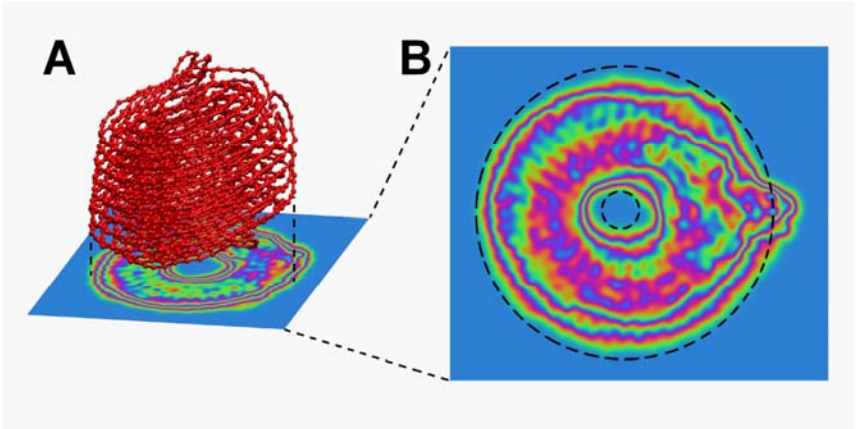

Fig. S4 Measurement of toroid dimensions. First, the toroid is projected to a density map, on a plane perpendicular to the axes through the hole, as shown in (**A**). Then the projected density map was measured using two circles, one to fit the outer perimeter and one to fit the hole (**B**). Finally, the toroid diameter is just the arithmetic average of the diameters of these two circles. Toroid thickness is the difference of the radii of these two circles.