bioRxiv preprint doi: https://doi.org/10.1101/375626; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Hierarchical multiscale simulation of DNA condensation

Short title: Multiscale simulation of DNA condensation

Tiedong Sun,[1] Alexander Mirzoev,[1] Vishal Minhas[1], Nikolay Korolev,[1] Alexander P. Lyubartsev,[2]*

Lars Nordenskiöld[1]*

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

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

*Corresponding authors, E-mail:

larsnor@ntu.edu.sg (L.N.); alexander.lyubartsev@mmk.su.se (A.P.L.)

bioRxiv preprint doi: https://doi.org/10.1101/375626; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

DNA condensation is of great importance for DNA packing *in vivo* with applications in medicine, biotechnology and polymer physics. Rigorous modeling of this process in all-atom MD simulations is presently impossible due to size and time scale limitations. We present a hierarchical approach for systematic multiscale coarse-grained (CG) simulations of DNA condensation induced by the three-valent cobalt(III)-hexammine ($CoHex^{3+}$). We extract solvent-mediated effective potentials for a CG model of DNA and demonstrate aggregation of small (100 bp) DNA in presence of explicit $CoHex^{3+}$ ions to a liquid crystalline type hexagonally ordered phase. Following further coarse-graining, simulations at mesoscale level are performed. Modeling of a 10 kbp-long DNA molecule results in formation of a toroid with distinct hexagonal packing in agreement with Cryo-EM observations. The procedure is based on the underlying all-atom force field description and uses no further adjustable parameters and may be generalized to modeling chromatin up to chromosome size.

bioRxiv preprint doi: https://doi.org/10.1101/375626; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

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 would adopt a random coil conformation expanding over 100 μm, 46 such DNA molecules are packed inside the confined space of about 10 μm in the human cell nucleus. Similarly, in sperm heads, bacteria and viruses, DNA is extremely densely condensed, being 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 fibers, 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 has been vastly studied experimentally with a variety of techniques such as X-ray diffraction (19), Cryo-EM (7, 18) and more recently with single molecule techniques (20). 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 the condensation of DNA induced by multivalent cations resulting in the formation of the ordered DNA liquid crystalline phase (for short molecules) and the collapse of DNA into toroidal structures (for kbp long molecules). Specifically, there is a lack of rigorous theoretical modeling

bioRxiv preprint doi: https://doi.org/10.1101/375626; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

approaches that are able to predict and reproduce these phenomena from basic physico-chemical 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 that result in a correlated fluctuation in the instantaneous positions of the condensed counterions on DNA, leading to a net attractive force between DNA molecules (reviewed in several works (2, 21)). In the case of flexible multivalent cations like the polyamines or oligopeptides, the attraction is also generated by the "bridging" effect (22). While the origin of multivalent ion induced attraction between aligned DNA-DNA molecules is clear and well described by different polyelectrolyte theories, the spontaneous transition of short DNA molecules to a hexagonal liquid crystalline phase and the formation of toroids from kbp long DNA molecules, have been less well described theoretically. These phenomena, being determined by electrostatic forces and by DNA mechanical properties, are important and biologically relevant examples of polyelectrolyte self-assembly and collapse.

Computer modeling of DNA condensation to an ordered phase or of single polymer collapse to bundles and toroids have with few exceptions been performed with a description of DNA as a chain of beads using parameterized harmonic bonds. The bending flexibility was commonly tuned to reproduce the DNA mechanical properties from persistence length data (23-26). Furthermore, the DNA-DNA attraction was generally modelled by empirical potentials and only in a few of these works electrostatic effects were treated directly by including explicit balancing counterions, however, without added salt (23-25). Depending on the choice of parameter values these simulations demonstrated formation of bundles, coils and toroids, with the presence of three-valent ions (strong electrostatic coupling) and stiff polymers favoring toroid formation. The

bioRxiv preprint doi: https://doi.org/10.1101/375626; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

mechanism of toroid formation was found to follow a complex pathway generally involving the formation of a nucleation loop followed by a growth process (25). Stevens investigated bundle formation of short generic polyelectrolytes of some relevance to the formation of a liquid a crystalline phase from short persistence length DNA molecules (23). Bundle formation of aligned polymer chains was observed.

Common to all these approaches is that they treat generic polymer molecules without explicit presence of added salt using 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 were predicted, the hexagonal toroid arrangement was only recently observed (26). Furthermore, the experimentally important phenomenon of the formation of hexagonally ordered liquid crystalline phase of short DNA molecule in the presence of multivalent counterions has not been theoretically demonstrated.

In recent years advances in computer 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 (27, 28). Yoo and Aksimentiev developed improved ion-phosphate interaction force field parameters and performed all-atom MD simulations of an array of 64 parallel duplex DNA. They demonstrated the correct hexagonal packing of DNA , which was absent in simulation with standard CHARMM or AMBER force field parameters (29). The same authors also investigated the physical mechanism of multivalent ion-mediated DNA condensation at atomistic level. The results supported a model of condensation driven by entropy gain due to release of monovalent ions and by bridging cations (27). However, all-atom MD simulation of DNA liquid crystalline formation or kbp-size DNA toroid condensation is presently not computationally feasible and hence multiscale approaches linking atomistic and coarse-grained (CG) levels of description are necessary (30).

bioRxiv preprint doi: https://doi.org/10.1101/375626; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Within a systematic bottom-up multiscale modeling scheme, the macromolecules are reduced to a CG description with effective sites representing groups of atoms (31). Typically, the term ”systematic coarse-graining” in molecular modeling 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) in terms of CG sites and their bonding. This results in a reduction of the number of degrees of freedom in the system. The following step in coarse-grained modeling is to determine the effective potentials. A number of coarse-graining methods have been developed to achieve this with respective strengths and weaknesses. Here we use structure-based coarse-graining, following the inverse Monte Carlo (IMC) procedure (32).

Although some bottom-up approaches have recently emerged (reviewed in (30)), they usually did not treat both electrostatics and solvent effects rigorously or did not reach the mesoscopic scale of DNA condensation. An alternative to the systematic bottom-up CG approach for modeling DNA at mesoscale level is represented by recent work by the de Pablo group. This approach is based on the 3-Site-Per-Nucleotide (3SPN) DNA CG model, which is parameterized in spirit similar to all-atom force fields, using a top-down approach to fit the model parameters to experimental data (DNA thermal denaturation) (33). This CG DNA model was recently combined with explicit ion-phosphate potentials obtained in a bottom-up approach based on underlying atomistic simulations (34). The DNA model combined with explicit ions was used in simulations of 4 kbp DNA packaging in the presence of multivalent ions inside a virus capsid (35). The 3SPN model without explicit ions was also used in simulations of nucleosome sliding (36).

The DNA condensation to a hexagonally ordered phase as well as DNA toroid formation, both induced by the presence of multivalent cations, are phenomena clearly intrinsic to the DNA molecule and inherent in its physico-chemical properties. We hypothesize that this behavior can be predicted by a bottom-up approach that is based on state-of-the-art all-atom molecular

bioRxiv preprint doi: https://doi.org/10.1101/375626; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

dynamics (MD) simulations, which is followed by structure-based CG simulations with IMC derived effective interaction potentials for the CG model without further adjustable parameters. Here, we perform systematic multiscale structure-based CG simulations of DNA with explicit electrostatic interactions included, starting from all-atom description going up to mesoscale modeling of DNA. We use a substantially improved CG DNA model with a topology similar to that introduced in (37, 38) with 5 beads per two-bp unit and explicit mobile ions. The systematic coarse-graining follows the IMC approach to extract solvent mediated effective CG potentials for all interactions in the system. The model with the IMC-developed potentials is validated in CG simulations of DNA persistence length as a function of monovalent salt. We demonstrate DNA condensation induced by the three-valent $CoHex^{3+}$ ion for short DNA duplexes resulting in a bundled phase with hexagonal ordering. Furthermore, adopting a second level "super coarse-grained" DNA beads-on-a-string model, we show that the approach predicts a hexagonally ordered liquid crystalline-like phase of short DNA and toroid formation in hexagonal arrangement for kbp-size long DNA. In order to further advance our knowledge and pave the ground for detailed analysis of the DNA condensation within various relevant contexts such as compaction of DNA in chromatin, it is important to develop chemically informed DNA models that do not rest on adjustable parameters and that have predictive power to be trusted in modelling phenomena at mesoscale level in experimentally unexplored scenarios. The present approach represents a first and successful step in this direction.

Results and Discussion

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 (39), we perform all-atom MD simulations for a system consisting of four DNA oligonucleotides in the presence of $CoHex^{3+}$ (Fig 1E). We then extract

bioRxiv preprint doi: https://doi.org/10.1101/375626; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

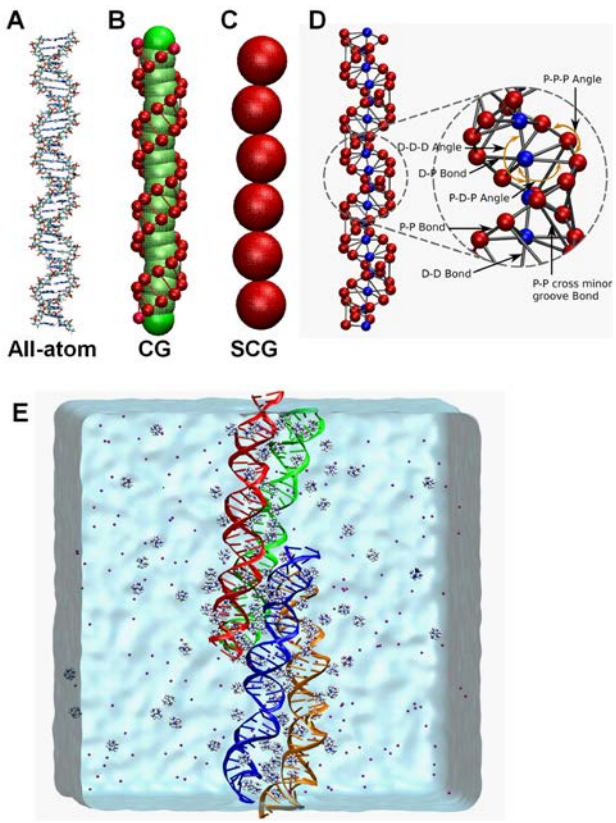

Fig. 1. (**A-D**) DNA models at multiple scales and (**E**) a snapshot of all-atom MD simulations of DNA condensation. (**A**) The all-atom DNA model. (**B**) The first level of DNA coarse-graining with two bead types. (**C**) The SCG model is built upon the CG model with one bead type. (**D**) The detailed presentation of the CG DNA model in (**B**). Bond interactions, D-P, P-P (backbone), D-D and P-P (cross minor groove) are shown in grey. Angle interactions (P-D-P, P-P-P and D-D-D) are indicated by orange arrows. (**E**). The MD simulation in a cubic box of 15 nm size with four 36 bp DNA double helices (colored blue, red, green and orange), ions (140 $CoHex^{3+}$, 140 K^{+}, 95 Na^{+}, 375 Cl^{-} shown as spheres); and explicit water (shown as blue surface). The snapshot demonstrates formation of a bundle consisting of four DNA molecules.

rigorous solvent mediated effective potentials for the CG description of DNA using the IMC method. The potentials are constructed to reproduce selected average structural properties of the fine-grained system such as radial distributions functions (RDF) and bond length and angle distributions. The present CG model of DNA (Fig. 1B) is detailed in Fig. 1D. The topology is the same as our previous model (37, 38) but with all interaction potentials derived by the IMC method from the underlying atomistic simulations. These potentials are tabulated and contain no adjustable parameters. Particularly, all non-bonded interactions are modeled by potentials which are capable of describing the short-range distribution of ions around DNA in solution and implicitly contain the effects of the solvent water. The details are given in the “Methods” section below. Subsequently, we use these potentials to perform CG MD simulations 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 condensation at mesoscale level. We perform simulations for a

bioRxiv preprint doi: https://doi.org/10.1101/375626; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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, we first perform 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. 1E. We perform three independent 1 μs-long simulations that are subsequently 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 (39), the system shows DNA-DNA attraction and aggregation of DNA into fiber-like bundles induced by $CoHex^{3+}$ as shown in Fig. 1E, representing a snapshot at the end of one of the three trajectories. DNA fibers are formed across the periodic boundaries in all three independent simulations (see Supplementary Fig. S1, A-C). The snapshots show some variability in the character of DNA-DNA contacts over the periodic images. Although all three simulations are technically converged (by standard criteria such as constant total energy and their components), when aggregation occurs the DNA bundles may be trapped in metastable states, which is unavoidable in any all-atom MD simulation of a complex system. This is why we perform three independent simulations and used the average RDFs. Since CG models generally present a smoother free energy surface than their fine-grained counterpart, more efficient sampling can be achieved with CG models. After obtaining effective potentials of the CG model, longer MD simulations of the same system with four DNA molecules, are performed (see more details in the next section). A final snapshot from one CG-MD simulation of this four DNA system is illustrated in the Supplementary Fig S1D. In the CG MD simulations a straight fiber configuration is easily reached. This result is stable and does not vary in different simulations.

bioRxiv preprint doi: https://doi.org/10.1101/375626; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Validation of the CG DNA model and of the bottom up IMC approach. First we validate the approach and the CG DNA model (described above) against experimental persistence length data as a function of monovalent salt. We perform all-atom MD simulation of a single 40 bp DNA in the presence of physiological salt (130 mM NaCl). The all-atom trajectory is run for 2 μs and demonstrates stability of helical parameters and the elasticity of the DNA double helix (data not shown). We then extract effective IMC potentials (see next section and Methods for details), following which, we proceed to run several CG simulations of a single 500 bp long DNA molecule in the presence of varying concentration of monovalent salt in the range 0.1 to 100 mM. The dependence of persistence length as a function of salt was calculated and compared with experimental data (40-46) (Fig. 2; for details of persistence length calculation see Methods section). We include experimental persistence length data from several sources as there is a large variation in experimental results depending on the method used and the procedure for analyzing the original measurements. For comparison, we also include values predicted by other CG DNA models with explicit ions, obtained using bottom-up approaches based on underlying all-atom MD simulations (34, 47, 48). The present model demonstrates very good agreement with experiments, given the variation in experimental data. Generally the performance of our model is as good as (or better than) other available explicit ion CG DNA models. This accurate prediction of the effect of electrostatic interactions on DNA flexibility lends confidence to the present

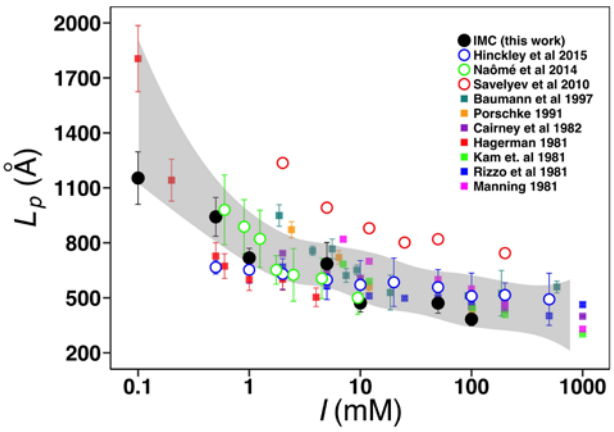

Fig. 2. Dependence of DNA persistence length (L_p) on NaCl concentration (I). Solid black circles represent this work. Hollow circles denote results from other explicit ion modeling data: blue (34), green (47), red (48). Solid squares are experimental results: dark cyan (40), orange (41), dark magenta (42), red (43), green (44), blue (45), magenta (46). Shaded area is the spline approximation of the experimental data with confidence level 0.99995.

bioRxiv preprint doi: https://doi.org/10.1101/375626; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

approach. The results shows that the model well represents both DNA intrinsic flexibility and electrostatic interactions. A more detailed discussion of DNA flexibility and persistence length prediction from the CG model and how the DNA flexibility depends on choice of underlying force field will be presented elsewhere (Minhas *et al*., to be published).

Building a Coarse-Grained DNA model with effective solvent mediated potentials obtained from IMC. Following the all-atom MD simulations of the system with four DNA molecules in all-atom description containing explicit ions ($CoHex^{3+}$, Na^{+}, K^{+} and Cl^{-}) and water as described above, we proceed to extract the effective potentials for the CG model based on the mapped RDFs using the IMC procedure. The trajectories generated by the MD simulations are mapped from the all-atom to the CG representation and then and we use them 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. Examples of calculated RDFs are shown in Fig. 3A-C. Following the IMC method, all effective interaction potentials for the CG DNA model are derived simultaneously, so that all correlations between different interaction terms are accounted for. All RDFs and effective potentials for the CG DNA model in presence of $CoHex^{3+}$ can be found in Supplementary Fig S2. Supplementary Fig. S3 illustrates convergence in the IMC calculations.

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 are replaced by the same number of Mg^{2+} ions (with the corresponding decrease in number of Cl^{-} counterions), resulting in the absence of DNA condensation, which is in agreement with the experimentally known fact that the divalent Mg^{2+} does not induce DNA condensation (1) (data not shown). We follow the same modeling protocol as for the case of $CoHex^{3+}$ to derive effective

bioRxiv preprint doi: https://doi.org/10.1101/375626; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

potentials for this control model. Selected distribution functions and effective potentials are plotted in Figs. 3A-F.

It may be noted that the D-D RDFs in the Mg^{2+} and $CoHex^{3+}$ systems between the central beads of different DNA molecules (Fig. 3A) appear very different. However, the final effective potentials for the D-D pair from both simulations (Fig. 3D) 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. In the Mg^{2+} system, the RDF at distances below 25 Å has low amplitude since different DNA fragments repel each other due to electrostatic interactions. The presence of Mg^{2+} counterions cannot overcome this repulsion. In the more strongly coupled $CoHex^{3+}$ system, the presence of the trivalent counterions creates an overall attraction between DNA segments that results in DNA condensation. Therefore, it is expected that similar D-D

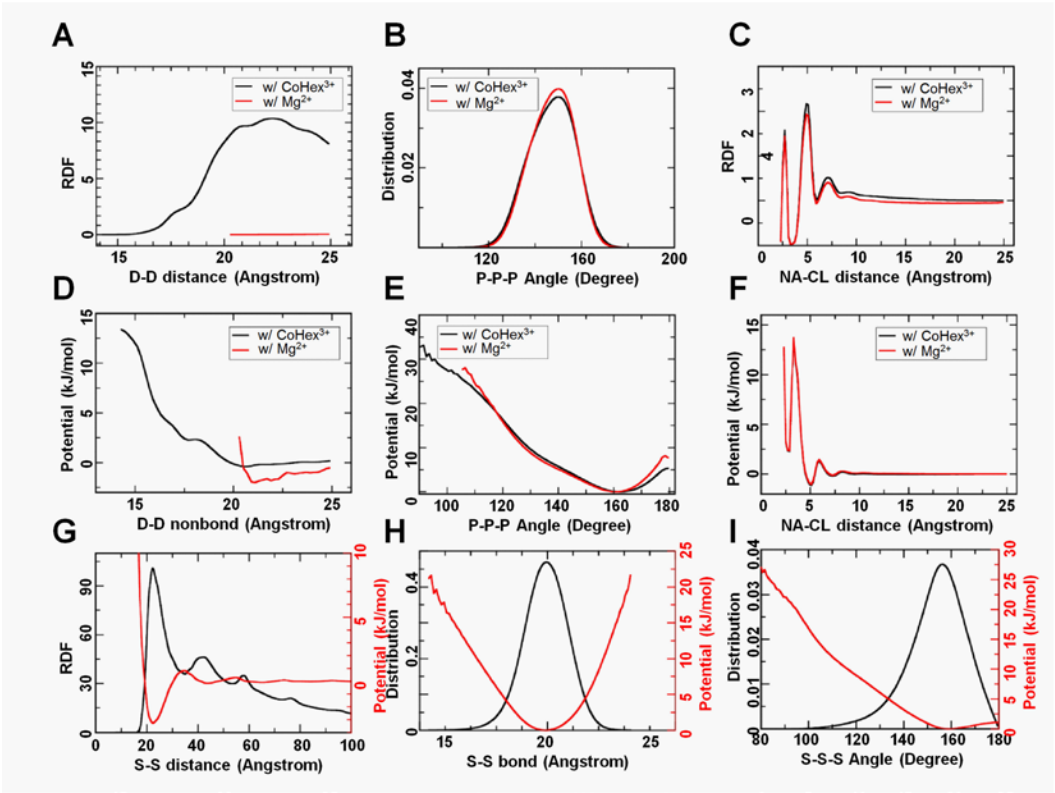

Fig. 3. Selected RDFs (top row), effective potentials of the CG model (middle row) and all RDFs and effective potentials of the SCG model (bottom row). In (**A-F**), three distributions and potentials are presented: D-D non-bonded pair, P-P-P angle along backbone and Na-Cl non-bonded pair, from both 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**).

bioRxiv preprint doi: https://doi.org/10.1101/375626; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

potentials result in different D-D RDFs if different ions are present in the system.

Furthermore, the effective potentials for monovalent ion-ion interactions obtained under different conditions are virtually the same as illustrated in Fig. 3F for the case of the Na-Cl pair. In Fig. 3C the corresponding Na-Cl RDF shows, however, a small but noticeable difference between the aggregating ($CoHex^{3+}$-system) and non-aggregating (Mg^{2+}-system) simulations due to different average DNA-DNA distances. However, the final effective potentials from both systems are identical (Fig. 3F). We attribute this to the fact that correlations between the different interaction terms are well represented in the present model. In fact, all ionic potentials for the same interaction types, but extracted from different underlying all-atom MD simulations for the present CG DNA model, are indistinguishable or very similar. As a further illustration, Supplementary Fig. S4 compares ionic potentials extracted from three DNA all-atom MD simulations having different ionic and different DNA compositions. This behavior implies good transferability of the derived potentials (see below for further discussion).

DNA aggregation in the CG simulations. Having rigorously extracted effective potentials for the CG DNA model on the basis of the all-atom simulations of the system with four DNA molecules in the presence of $CoHex^{3+}$, 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 non-bonded interaction (shown in the figures, e.g. Fig. 3) within the cut-off distance, determined by the IMC procedure. This treatment of the long-range electrostatic interactions was validated in our previous work (49). 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.

bioRxiv preprint doi: https://doi.org/10.1101/375626; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

After the significant reduction in the number of degrees of freedom by the coarse-graining, we can easily simulate DNA condensation in a 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 of 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.

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 bundles together. The DNA condensate particle gets larger, forming a single fiber-like particle at the end of the simulation. Remarkably, the DNA molecules demonstrate short-ranged hexagonal arrangement in the fiber bundle.

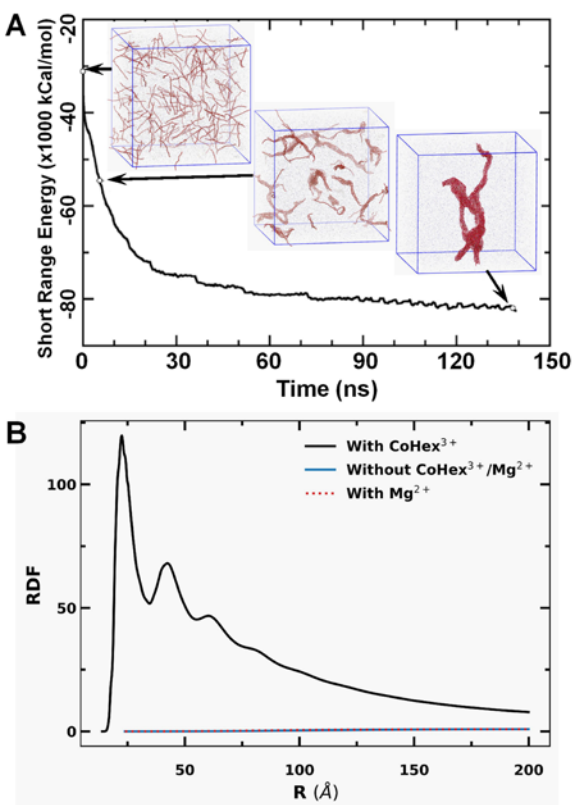

Fig. 4 (**A**) Snapshots and short-range interaction energy profile from CG DNA simulation. Time points corresponding to each snapshot are indicated with black circles on the energy curve. DNA molecules are rendered red. (**B**) D-D RDF comparison between the CG simulations with the presence of Mg^{2+} ions (red dashed), $CoHex^{3+}$ (black) and monovalent salt only (blue).

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, 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

bioRxiv preprint doi: https://doi.org/10.1101/375626; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

24 Å – 27 Å (assuming hexagonal ($r = 4\pi/q\sqrt{3}$) or lamellar ($r = 2\pi/q$) packing). A shorter separation observed for bundled DNA in simulations compared to experiment is likely due to the CHARMM force field used in the underlying all-atom MD simulations (50).

To test the robustness of the CG DNA model, we conduct another simulation where the $CoHex^{3+}$ ions are replaced by K^+ and Na^+ ions with equivalent amount of charge, 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 (blue line). In contrast to the system with $CoHex^{3+}$, the amplitude of the D-D RDF is low in value, suggesting a large distance between DNA molecules. Additionally, Fig. 4B displays the D-D RDF (dashed red line) obtained from a CG simulation where the $CoHex^{3+}$ ions are replaced by Mg^{2+} ions, with all potentials obtained from the all-atom system comprising four DNA molecules and Mg^{2+} ions, as mentioned above in relation to Fig 3D-F. In agreement with the experimental data (1), this system is repulsive and phase separation does not occur.

Hence, we can conclude that our CG DNA model is robust and produces realistic DNA aggregation behavior in large-scale simulations. The result illustrated in Fig. 4B also lends support for the transferability of the effective ionic potentials obtained with the present approach. The CG simulation resulting in the red (repulsive) D-D RDF are performed for a system with 200 DNA molecules of 100 bp length containing only monovalent ions. In such a system the expected macroscopic behavior is that DNA condensation should not occur. However, the effective ionic potentials used in the CG simulation resulting in this non-aggregating system are obtained from an all-atom MD simulation containing $CoHex^{3+}$ ions that displays aggregation and bundling at all atom level (Fig. 1E) and in the CG simulations (Fig. 4).

As a further test of the transferability of the ionic potentials, we compare effective potentials for all monovalent ion interactions obtained by the IMC procedure from two different underlying

bioRxiv preprint doi: https://doi.org/10.1101/375626; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

all-atom MD simulations. These all-atom systems have the presence of $CoHex^{3+}$ or Mg^{2+} (mentioned above in relation to Fig. 3) with one displaying attraction ($CoHex^{3+}$-system) and the other displaying repulsion (Mg^{2+}-system). CG simulations for two systems containing 25 CG DNA molecules in the presence of $CoHex^{3+}$ are conducted. One system has monovalent ion effective potentials originating from the repulsive all-atom MD Mg^{2+}-system. The other system has monovalent ion potentials originating from the attractive all-atom MD $CoHex^{3+}$-system. All other potentials (DNA and $CoHex^{3+}$) are taken from the $CoHex^{3+}$-system. The D–D RDF displayed in Fig. S5 illustrates that the macroscopic behavior in these two simulations, is very similar resulting in attraction and phase separation in spite of the fact that the monovalent ion potentials are extracted from two separate all-atom MD simulations that exhibit different macroscopic behaviour.

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-grained (SCG) model (Fig. 1C). The excellent behavior of the CG DNA model allowed us to confidently build a DNA model with even lower resolution and faster 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, as well as all electrostatic interactions were treated implicitly by the effective potentials 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

bioRxiv preprint doi: https://doi.org/10.1101/375626; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

distributions. These two terms may in principle be modeled 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 or Debye-Hückel 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 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 a 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 perform simulations for a single 10 kb long DNA molecule with the SCG DNA model. The simulations start from a fully extended DNA conformation. After a short relaxation at the beginning, a loop or a bundle can form at either end of the DNA (see Supplementary Movie S1). The first snapshot in Fig. 5C shows the DNA conformation after such a loop formation.

bioRxiv preprint doi: https://doi.org/10.1101/375626; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

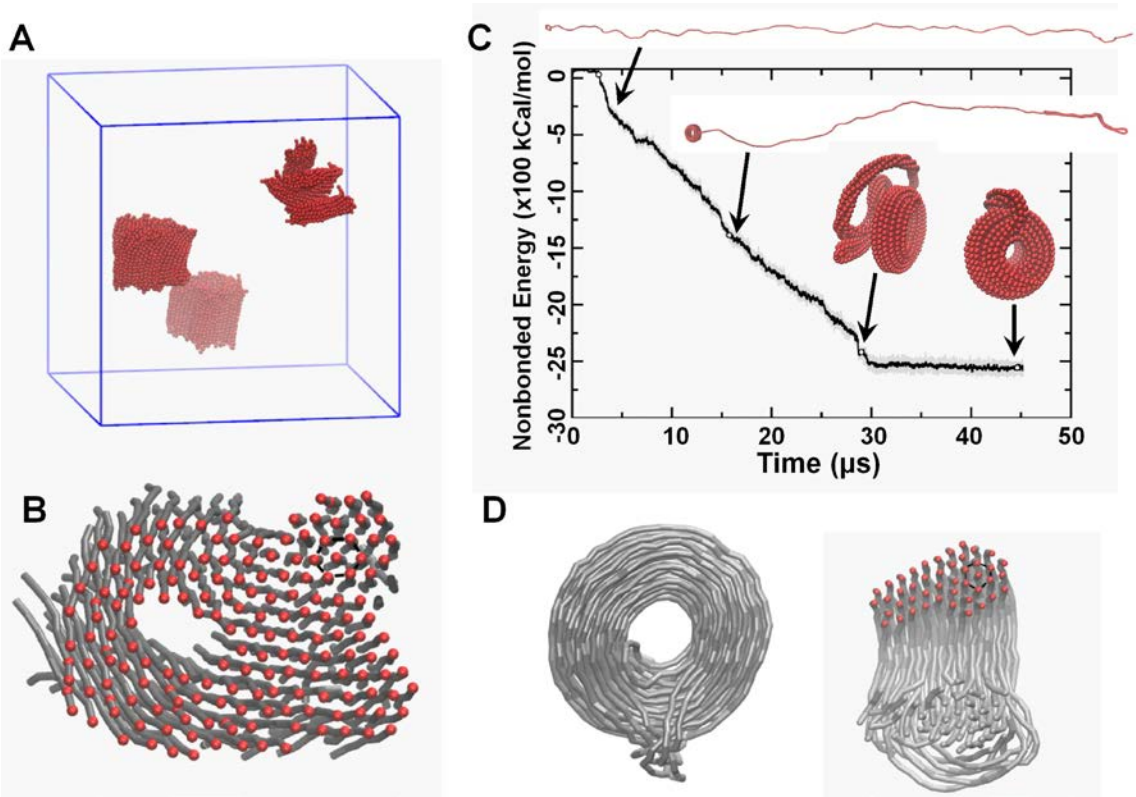

Fig. 5. DNA aggregation simulated by the SCG DNA model. (**A**) Final configuration of DNA aggregation in a simulation with 400 DNA molecules. (**B**) Cross-section of one of the DNA condensed particle shown in (**A**). (**C-D**) Formation of toroidal structures in 10 kbp DNA SCG simulations. (**C**) Energy profile and snapshots from one of the simulations. (**D**) Structure of the DNA toroid. Cartoon on the right-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.

Subsequently, the loop and bundle play the role of nucleation sites, 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 Supplementary Movie S3). The non-bonded interaction energy decreases as 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 shows that DNA is organized in hexagonal arrangement. These structural features are consistent with the reported electron microscopy studies (5).

Analyzing multiple simulations of single DNA reveals that toroids are mainly formed in two ways. The first scenario suggests that a single loop at one end initiates the toroid formation, while the other end may eventually form a fiber 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

bioRxiv preprint doi: https://doi.org/10.1101/375626; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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. Often, toroids are not formed but the final condensed structure is fiber-like, which has also been observed in EM experiments (18). In the simulations, the toroid shape is, however, somewhat more frequently occurring. We also simulate a lambda-DNA size single DNA molecules (48 kbp) that also form toroids, here by the mechanism with one nucleation loop at both ends (data not shown).

Next we perform an analysis of the mechanism of DNA condensation that results in toroid or fiber bundle formation. Here we pay specific attention to the pathways and the initial events during DNA condensation that leads to formation of the major final states of toroid and fiber-like morphologies. To this end we perform 67 independent SCG simulations of the 10 kbp DNA (having same initial coordinates but with different starting velocities). Although the mechanism of DNA condensation to either toroid or fiber is complex and stochastic, it is possible to identify several intermediate states as well as transitions between them. The formation of either toroid or fiber can be divided into two stages. The first stage can be characterized as nucleation. Secondly, the growth of the nucleation site occurs by pulling in more and more DNA chains, and by the sliding mechanism mentioned above. Fig. 6 summarizes the early event intermediate states and transitions between them. Supplementary Table S1 gives a description of those frequent initial transient structures. In the first few nanoseconds of each simulation, several transient states can be observed, such as state **b** and state **k** in the figure. These are usually simple structures with short lifetimes. As the simulation proceeds, more stable structures can be formed, for example, bundles consisting of three DNA chains (state **c** in the figure) and the double DNA chain loop (state **g**).

bioRxiv preprint doi: https://doi.org/10.1101/375626; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Based on these more stable structures, the "growth" mechanism can be effective with the remaining part of the DNA chain being pulled in to form larger and more stable structure.

In the early stage of the simulation, fiber-like structures and toroid-like structures can convert between each other. These structures are usually no more than three DNA chains thick. Hence, the energy barrier involved in these transitions should not be prohibitive. For example, a three DNA chain bundle (state **c**) can convert to a single DNA chain loop (state **h**) to form a toroid-like structure (state **i**). On the other hand, this toroid-like structure (state **i**) is soft enough, so that it can go to a bundle (state **d**) just by closing the hole in the middle. These transitions between fiber-like and toroid-like states rarely happen in the subsequent stages of the simulation when more DNA chains become involved in the structures.

Finally, we measure the dimensions of toroids formed by the 10kb DNA in multiple simulations shown in Fig. 6 insert (with details of calculations given in Supplementary Fig S6). 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

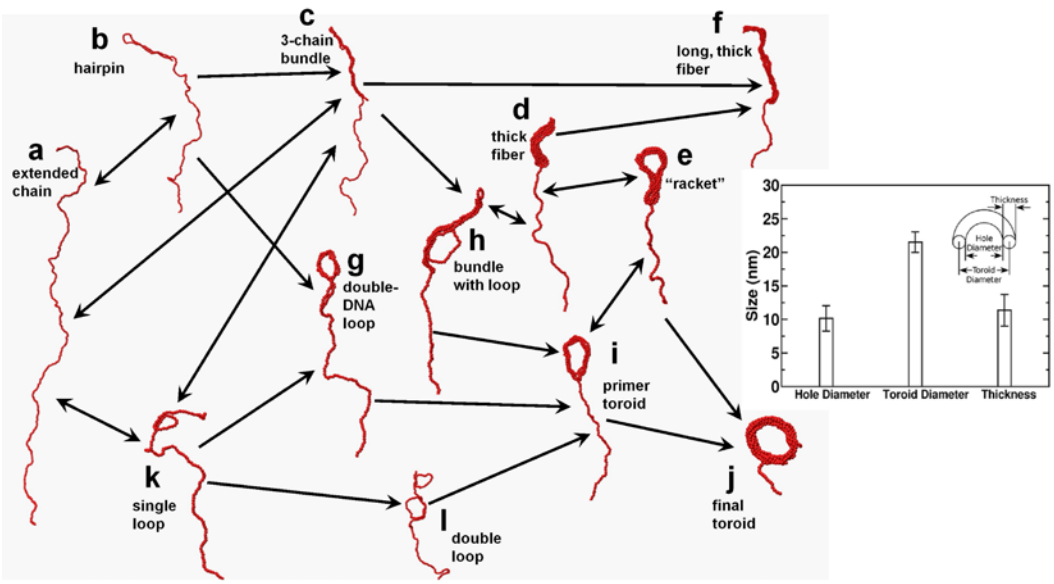

Fig. 6. Intermediate states during formation of DNA toroid and fiber. The structural features of each state are described in more details in Supplementary Table S1. Arrows indicate the transition directions among the states, as observed in the SCG simulations. Insert on the right shows statistics of 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. S6 for details.

bioRxiv preprint doi: https://doi.org/10.1101/375626; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

also substantially smaller than the dimensions observed in the experiments (18). These values are very reasonable given the differences between experiments and simulations. The reason for smaller diameter and thickness in simulations as compared to experiments, 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, on the other hand, is 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.

Conclusions

In conclusion, we have developed a rigorous hierarchical multiscale simulation scheme, which enables simulation of DNA condensation at mesoscale levels. The phenomenon of multivalent ions 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

bioRxiv preprint doi: https://doi.org/10.1101/375626; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 used no 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 (39). 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 (30, 33). It may also be noted that the present CG DNA model is not sequence specific, but such an extension can be implemented in the model (38).

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 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 fibers 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).

bioRxiv preprint doi: https://doi.org/10.1101/375626; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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.

Methods

All-atom Molecular Dynamics simulations. In the present work, we use an atomistic force field model as our high resolution reference in the IMC procedure to extract effective potentials. Specifically, the coarse-graining is started with all-atom molecular dynamics simulations using the CHARMM27 force field (51). In previous work (39) 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) preservation of DNA B-form internal structure and its excellent agreement with experimental DNA persistence length data (38), see above. The all-atom MD simulation is set up with four double helix DNA molecules, with four 36 bp-long double helix DNA molecules placed in a cubic simulation cell with periodic boundaries. The DNA sequence is the same as in our previous work (39) 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, modelled as in our previous study are present to induce DNA condensation (39). 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 (29) are used throughout all simulations for all ions except $CoHex^{3+}$ and the TIP3P water model is utilized. The system is illustrated in Fig. 1E.

bioRxiv preprint doi: https://doi.org/10.1101/375626; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 implemented in GROMACS 5.1 (52), 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. The velocity rescale thermostat and Parrinello-Rahman barostat are adopted to regulate temperature and pressure respectively. Electrostatic interaction is treated with particle mesh Ewald 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 the same number of 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 $CoHex^{3+}$. For validation of the CG DNA model by persistence length calculation, a single 40 base-pair DNA with sequence 5′-GGATTAATGGAACGTAGCATATT-CTTCAAGTTGTCACGCC (42.5% GC content) was used. To mimic physiological conditions the NaCl concentration is set to 130 mM. All other conditions are similar to the aforementioned simulations with $CoHex^{3+}$.

Coarse-grained DNA models. In order to reach the mesoscale level of DNA, we have performed coarse-graining at two spatial scales 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

bioRxiv preprint doi: https://doi.org/10.1101/375626; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(SCG) model (Fig. 1C). The CG DNA model is designed according to the same concepts as our previous model, which was shown to well reproduce DNA persistence length dependence on salt concentration over a vast concentration range (38). It is simple enough in its design and yet captures the structural form and properties of double helical B-form DNA. Here, 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 (denoted “P”), while the other one (denoted “D”) represents the four nucleosides in between. 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). These pairwise bonds are D-D, D-P, P-P along the strand and P-P cross-strand over the minor groove defined as a shortest P-P bond over the groove. The angle bonds are D-D-D, P-D-P, and P-P-P.

The model, which includes explicit mobile ions and explicit charges of DNA phosphate groups, pays particular emphasis to electrostatic interactions, where every P-bead has charge -1e, and D-beads are kept neutral. The solvent is considered implicitly, via use of the effective short-range potentials and screening of the electrostatic interactions according to the solvent dielectric permittivity.

Despite looking similar to its ancestor (38), the present model has several very important features being considerably different from what had been reported by us earlier. First the underlying all-atom MD simulations are significantly longer (an order of magnitude compared to the previous model). Secondly, it is fully based on effective potentials derived from the underlying all-atom MD simulation using the Inverse Monte Carlo method. Particularly, all non-bonded interactions are not simply a combination of electrostatic and repulsive truncated Lennard-Jones (as in the case of ref (38)), but are rather complex functions which are capable of

bioRxiv preprint doi: https://doi.org/10.1101/375626; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

describing distribution of ions around DNA in solution according to the underlying atomistic simulations**.** In the present model, both bonded and non-bonded potentials (shown in Fig. S2) are tabulated and are not fitted to a specific functional form and contain no adjustable parameters.

The SCG DNA model prioritizes computation performance over complexity, being 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 the solvent is considered implicitly, but electrostatic interactions are implicitly included into these effective potentials between the S-beads. The effective potentials (Fig. 3 G-I) are derived after mapping the CG-model trajectory to this SCG followed by IMC.

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. These are 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 parameters entering into this model and it rests solely on the all-atom CHARMM27 DNA 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

bioRxiv preprint doi: https://doi.org/10.1101/375626; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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. Each final set of RDF is calculated as an average over the equilibrated sections of all three independent trajectories. The total length of the equilibrated trajectories used is equal to 1.5 μs.

The IMC inversion calculation is carried out with the MagiC software (53) (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 as trial 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 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. The procedure described above is followed in a similar way for deriving the effective CG potentials from the other all-atom systems, namely the one with four DNA and Mg^{2+} ions and the one comprising a single DNA and monovalent salt.

Coarse-grained simulations. The tabulated interaction potentials for the CG DNA model, obtained as described above are used in the CG MD simulations with a significantly bigger

bioRxiv preprint doi: https://doi.org/10.1101/375626; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 (54) 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. This procedure for CG simulations is conducted analogously for the systems with Mg^{2+} ions and with a single DNA at varying monovalent salt concentrations (persistence length validation).

Persistence length calculation. To validate the CG DNA model we test its performance in prediction of the experimental salt dependence of the DNA persistence length. We perform all-atom MD simulations of a single 40 bp DNA in the presence of physiological salt (130 mM NaCl). Following that we perform the mapping of this trajectory to the CG model and extract effective IMC potentials (Supplementary Fig. S7 shows all the effective potentials). A 500 bp long CG DNA molecule is simulated in the range of NaCl salt concentrations from 0.1 mM to 100 mM to estimate the dependence of the persistence length on ionic concentration. All the simulations are run for at least 3 μs and employ a similar protocol as we describe in the previous section. The persistence length calculated using different parts of the trajectory shows similar

bioRxiv preprint doi: https://doi.org/10.1101/375626; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

values, confirming the convergence of the simulations. These results are then compared with experimental results from similar concentrations. The persistence length is calculated using the formula

$$L_p = -\frac{L_c}{\ln(\langle cos\alpha \rangle)}$$

where L_c is the contour length of the DNA and $\langle cos\alpha \rangle$ is the average of the cosine between two adjacent segments over the length of the simulation for the corresponding contour length. Then $\ln(\langle cos\alpha \rangle)$ is plotted as a function of L_c. L_p is estimated by determining the slope of the plot, as done in our previous study (38). We divide our 500 bp DNA into 5 equal length segments. Data is collected with each DNA segment, which is 100 bp ($L_c \approx 330$ Å) long. That results in 5 data points under each ionic concentration. The standard deviation of L_p is evaluated with these 5 data points. Results are shown in Fig. 2.

Deriving effective potentials for the SCG model by Inverse Monte Carlo. The effective potential for the SCG model is derived in the same way as the CG model by the inverse Monte Carlo method. Specifically, the trajectories from MD simulation based on the CG model are mapped to a SCG trajectory by the mapping rules defined earlier. In this step, all species of ions are made implicit. With these trajectories as fine-grained reference, we have calculated one non-bonded RDF (S-S), one bond distribution (S-S) and one angle distribution (S-S-S), according to our definition to the SCG model. The cut-off for non-bonded interaction is set to be 200 Å. The inversion calculation is started with zero potential for non-bonded interaction, and with potential of mean force for bond and angle interactions, as similar to our previous calculation for CG model. Temperature and volume are kept constant during sampling with 100 parallel Monte Carlo simulations. The effective potential converges in 25 iterations.

bioRxiv preprint doi: https://doi.org/10.1101/375626; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Coarse-grained MD simulations using the SCG DNA model. The electrostatic interactions are treated implicitly in the SCG model as they are effectively included into the 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. First, velocity rescaling is used to regulate the temperature during the first 10^5 steps in order to stabilize the temperature to 298 K in the equilibration stage. Then the production stage of 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 the velocity rescaling. After that, 2×10^8 steps are performed in each simulation at the temperature of 298 K controlled 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.

Supplementary Materials

Supplementary material for this article is available at ????http://advances.sciencemag.org????

Fig. S1. Snapshots from different all-atom and CG MD simulations modelling DNA-DNA interaction in systems with $CoHex^{3+}$ cations.

Fig. S2. The complete set of RDF and converged effective potentials of the CG model.

Fig. S3. Convergence of the Inverse Monte Carlo procedure.

Fig. S4. Transferability of potentials from different systems I.

Fig. S5. Transferability of potentials from different systems II.

bioRxiv preprint doi: https://doi.org/10.1101/375626; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Fig. S6. Measurement of toroid dimensions.

Fig. S7. RDFs and potentials used for calculation of persistence length.

Table S1. Description of intermediate states during formation of DNA toroid and fiber.

Movie S1. Loop formation at one end of the DNA.

Movie S2. Growth of toroid.

Movie S3. DNA molecule forms single toroidal particle.

Movie S4. Growth of two toroids at the DNA ends

References and Notes

1. V. A. Bloomfield, DNA condensation by multivalent cations. *Biopolymers* **44**, 269-282 (1997).
2. W. M. Gelbart, R. F. Bruinsma, P. A. Pincus, V. A. Parsegian, DNA-inspired electrostatics. *Physics Today* **53**, 38-44 (2000).
3. P. Carrivain *et al.*, Electrostatics of DNA compaction in viruses, bacteria and eukaryotes: functional insights and evolutionary perspective. *Soft Matter* **8**, 9285-9301 (2012).
4. T. J. Thomas, T. Thomas, Collapse of DNA in packaging and cellular transport. *Int. J. Biol. Macromol.* **109**, 36-48 (2018).
5. N. V. Hud, K. H. Downing, Cryoelectron microscopy of lambda phage DNA condensates in vitreous ice: the fine structure of DNA toroids. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 14925-14930 (2001).
6. S. M. Klimenko, T. I. Tikchonenko, V. M. Andreev, Packing of DNA in the head of bacteriophage T2. *J. Mol. Biol.* **23**, 523-IN521 (1967).
7. A. Leforestier, F. Livolant, Structure of toroidal DNA collapsed inside the phage capsid. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 9157-9162 (2009).
8. J. Ausió, R. González-Romero, C. L. Woodcock, Comparative structure of vertebrate sperm chromatin. *J. Struct. Biol.* **188**, 142-155 (2014).
9. Y. M. Evdokimov, A. L. Platonov, A. S. Tikhonenko, Y. M. Varshavsky, A compact form of double-stranded DNA in solution. *FEBS Lett.* **23**, 180-184 (1972).
10. U. K. Laemmli, Characterization of DNA condensates induced by poly(ethylene oxide) and polylysine. *Proc. Natl. Acad. Sci. U.S.A.* **72**, 4288-4292 (1975).
11. L. C. Gosule, J. A. Schellman, Compact form of DNA induced by spermidine. *Nature* **259**, 333 (1976).
12. D. K. Chattoraj, L. C. Gosule, J. A. Schellman, DNA condensation with polyamines: II. Electron microscopic studies. *J. Mol. Biol.* **121**, 327-337 (1978).
13. R. W. Wilson, V. A. Bloomfield, Counterion-induced condensation of deoxyribonucleic acid. A light-scattering study. *Biochemistry* **18**, 2192-2196 (1979).
14. J. L. Sikorav, J. Pelta, F. Livolant, A liquid crystalline phase in spermidine-condensed DNA. *Biophys. J.* **67**, 1387-1392 (1994).
15. J. Pelta, F. Livolant, J. L. Sikorav, DNA aggregation induced by polyamines and cobalthexamine. *J. Biol. Chem.* **271**, 5656-5662 (1996).
16. M. Nakata *et al.*, End-to-end stacking and liquid crystal condensation of 6 to 20 base pair DNA duplexes. *Science* **318**, 1276-1279 (2007).

bioRxiv preprint doi: https://doi.org/10.1101/375626; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

17. C. C. Conwell, I. D. Vilfan, N. V. Hud, Controlling the size of nanoscale toroidal DNA condensates with static curvature and ionic strength. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 9296-9301 (2003).
18. N. V. Hud, I. D. Vilfan, Toroidal DNA condensates: Unraveling the fine structure and the role of nucleation in determining size. *Annu. Rev. Biophys. Biomol. Struct.* **34**, 295-318 (2005).
19. J. A. Schellman, N. Parthasarathy, X-ray diffraction studies on cation-collapsed DNA. *J. Mol. Biol.* **175**, 313-329 (1984).
20. L. R. Brewer, Deciphering the structure of DNA toroids. *Integr. Biol. (Camb).* **3**, 540-547 (2011).
21. N. Korolev, A. P. Lyubartsev, L. Nordenskiöld, Cation-induced polyelectrolyte-polyelectrolyte attraction in solutions of DNA and nucleosome core particles. *Adv. Colloid Interface Sci.* **158**, 32–47 (2010).
22. R. Podgornik, M. Licer, Polyelectrolyte bridging interactions between charged macromolecules. *Curr. Opin. Colloid Interface Sci.* **11**, 273–279 (2006).
23. M. J. Stevens, Bundle binding in polyelectrolyte solutions. *Phys. Rev. Lett.* **82**, 101-104 (1999).
24. M. J. Stevens, Simple simulations of DNA condensation. *Biophys. J.* **80**, 130-139 (2001).
25. Z. Ou, M. Muthukumar, Langevin dynamics of semiflexible polyelectrolytes: Rod-toroid–globule-coil structures and counterion distribution. *J. Chem. Phys.* **123**, 074905 (2005).
26. A. Dey, G. Reddy, Toroidal condensates by semiflexible polymer chains: Insights into nucleation, growth and packing defects. *J. Phys. Chem. B* **121**, 9291-9301 (2017).
27. J. Yoo, A. Aksimentiev, The structure and intermolecular forces of DNA condensates. *Nucleic Acids Res.* **44**, 2036-2046 (2016).
28. H. Ishida, H. Kono, H4 tails potentially produce the diversity in the orientation of two nucleosomes. *Biophys. J.* **113**, 978–990 (2017).
29. J. Yoo, A. Aksimentiev, 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 (2012).
30. P. D. Dans, J. Walther, H. Gomez, M. Orozco, Multiscale simulation of DNA. *Curr. Opin. Struct. Biol.* **37**, 29–45 (2016).
31. S. Takada, Coarse-grained molecular simulations of large biomolecules. *Curr. Opin. Struct. Biol.* **22**, 130-137 (2012).
32. A. P. Lyubartsev, A. Laaksonen, Calculation of effective interaction potentials from radial distribution functions: A reverse Monte Carlo approach. *Phys. Rev. E* **52**, 3730–3737 (1995).
33. D. M. Hinckley, G. S. Freeman, J. K. Whitmer, J. J. de Pablo, An experimentally-informed coarse-grained 3-site-per-nucleotide model of DNA: Structure, thermodynamics, and dynamics of hybridization. *J. Chem. Phys.* **139**, 144903 (2013).
34. D. M. Hinckley, J. J. de Pablo, Coarse-Grained ions for nucleic acid modeling. *J. Chem. Theory Comput.* **11**, 5436-5446 (2015).
35. A. Córdoba, D. M. Hinckley, J. Lequieu, J. J. de Pablo, A molecular view of the dynamics of dsDNA packing inside viral capsids in the presence of ions. *Biophys. J.* **112**, 1302-1315 (2017).
36. J. Lequieu, D. C. Schwartz, J. J. de Pablo, In silico evidence for sequence-dependent nucleosome sliding. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E9197-E9205 (2017).
37. Y. Fan, N. Korolev, A. P. Lyubartsev, L. Nordenskiöld, An advanced coarse-grained nucleosome core particle model for computer simulations of nucleosome-nucleosome interactions under varying ionic conditions. *PLoS One* **8**, e54228 (2013).

bioRxiv preprint doi: https://doi.org/10.1101/375626; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

38. N. Korolev, L. Di, A. P. Lyubartsev, L. Nordenskiöld, A coarse-grained DNA model parameterized from atomistic simulations by inverse Monte Carlo. *Polymers* **6**, 1655-1675 (2014).
39. T. Sun, A. Mirzoev, N. Korolev, A. P. Lyubartsev, L. Nordenskiöld, All-atom MD simulation of DNA condensation using ab initio derived force field parameters of cobalt(III)-hexammine. *J. Phys. Chem. B* **121**, 7761-7770 (2017).
40. C. G. Baumann, S. B. Smith, V. A. Bloomfield, C. Bustamante, Ionic effects on the elasticity of single DNA molecules. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 6185-6190 (1997).
41. D. Porschke, Persistence length and bending dynamics of DNA from electrooptical measurements at high salt concentrations. *Biophys. Chem.* **40**, 169-279 (1991).
42. K. L. Cairney, R. E. Harrington, Flow birefringence of T7 phage DNA: dependence on salt concentration. *Biopolymers* **21**, 923-934 (1982).
43. P. J. Hagerman, Flexibility of DNA. *Annu. Rev. Biophys. Biophys. Chem.* **17**, 265-286 (1988).
44. Z. Kam, N. Borochov, H. Eisenberg, Dependence of laser light scattering of DNA on NaCl concentration. *Biopolymers* **20**, 2671-2690 (1981).
45. V. Rizzo, J. Schellman, Flow dichroism of T7 DNA as a function of salt concentration. *Biopolymers* **20**, 2143–2163 (1981).
46. G. S. Manning, A procedure for extracting persistence lengths from light-scattering data on intermediate molecular weight DNA. *Biopolymers* **20**, 1751-1755 (1981).
47. A. Naômé, A. Laaksonen, D. P. Vercauteren, A solvent-mediated coarse-grained model of DNA derived with the systematic Newtion inversion method. *J. Chem. Theory Comput.* **10**, 3541–3549 (2014).
48. A. Savelyev, G. A. Papoian, Chemically accurate coarse graining of double-stranded DNA. *Proc.Natl.Acad.Sci.U.S.A.* **107**, 20340-20345 (2010).
49. A. A. Mirzoev, A. P. Lyubartsev, Effective solvent-mediated interaction potentials of Na+ and Cl- in aqueous solution: temperature dependence. *Phys. Chem. Chem. Phys.* **13**, 5722-5727 (2011).
50. J. Yoo, A. Aksimentiev, 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**, 8432-8449 (2018).
51. A. D. MacKerell, N. Banavali, 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 (2000).
52. M. J. Abraham *et al.*, GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19-25 (2015).
53. A. Mirzoev, A. P. Lyubartsev, MagiC: Software package for multiscale modeling. *J. Chem. Theory Comput.* **9**, 1512-1520 (2013).
54. S. Plimpton, Fast parallel algorithms for short-range Molecular Dynamics. *J. Comp. Phys.* **117**, 1-19 (1995).

Acknowledgments: General: We acknowledge the generous support of computer time allocation from the National Supercomputing Centre (NSCC) Singapore. We are indebted to Prof Aatto Laaksonen for discussions and suggestions. **Funding:** This work was supported by the Singapore Ministry of Education Academic Research Fund (AcRF) Tier 2 (MOE2014-T2-1-123

bioRxiv preprint doi: https://doi.org/10.1101/375626; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(ARC51/14)) and Tier 3 (MOE2012-T3-1-001) grants (to L.N.) and by the Swedish Research Council (to A.P.L.).

Author contributions: T.S., V.M. and A.M. performed simulations. T.S., V.M., 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.

Competing interests: The authors declare that they have no competing interests.

Data and materials availability: All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

bioRxiv preprint doi: https://doi.org/10.1101/375626; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Material

Hierarchical multiscale simulation of DNA condensation

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

bioRxiv preprint doi: https://doi.org/10.1101/375626; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1. Snapshots from different all-atom and CG MD simulations modelling DNA-DNA interaction in systems with $CoHex^{3+}$ cations.

Figure S1 (A-C) illustrates variability of DNA bundling from snapshots in the three all-atom MD simulations. The simpler energy profile and significantly longer sampling in the CG MD simulation results in minor variability between different simulations (Fig. S1D).

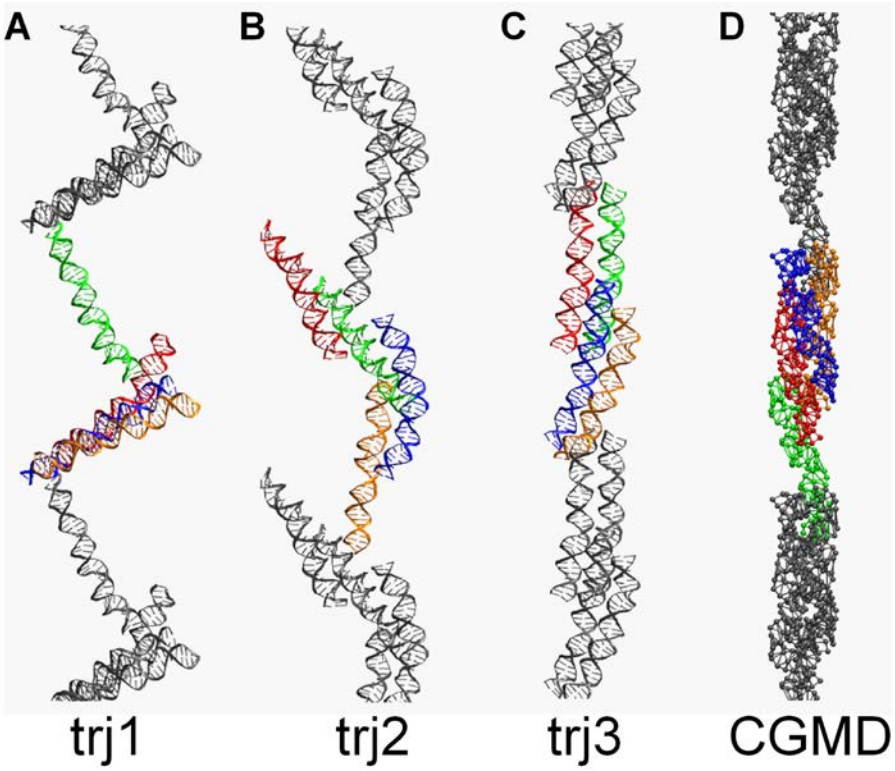

Fig. S1. (**A-C**) Final snapshot conformations obtained for all-atom trajectories in the MD simulations with $CoHex^{3+}$. DNA double helices are colored differently. Periodic images of the DNAs are colored grey. (**D**) A snapshot of the final conformation in a coarse grained MD simulation using conditions similar to the all-atom systems and interaction potentials determined from the simulations shown in (**A-C**).

bioRxiv preprint doi: https://doi.org/10.1101/375626; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

Below, all the RDFs and effective potentials corresponding to all interaction terms in the CG-model simulations with $CoHex^{3+}$ are displayed. DNA is modelled by four bead types (DT, PT, D, P) and 4 types of ions (CO, K, NA, CL) are also present in the system. The complete set of interaction parameters consists of 36 non-bonded terms, 4 bonded terms and 3 angle terms. Analysis of the plots allows observation of correlations among interaction terms. For example, there would be no potential minimum for the D-P non-bonded interaction at 6.5 Å if there is no correlation with other interactions, since there is no peak at 6.5 Å in the D-P RDF.

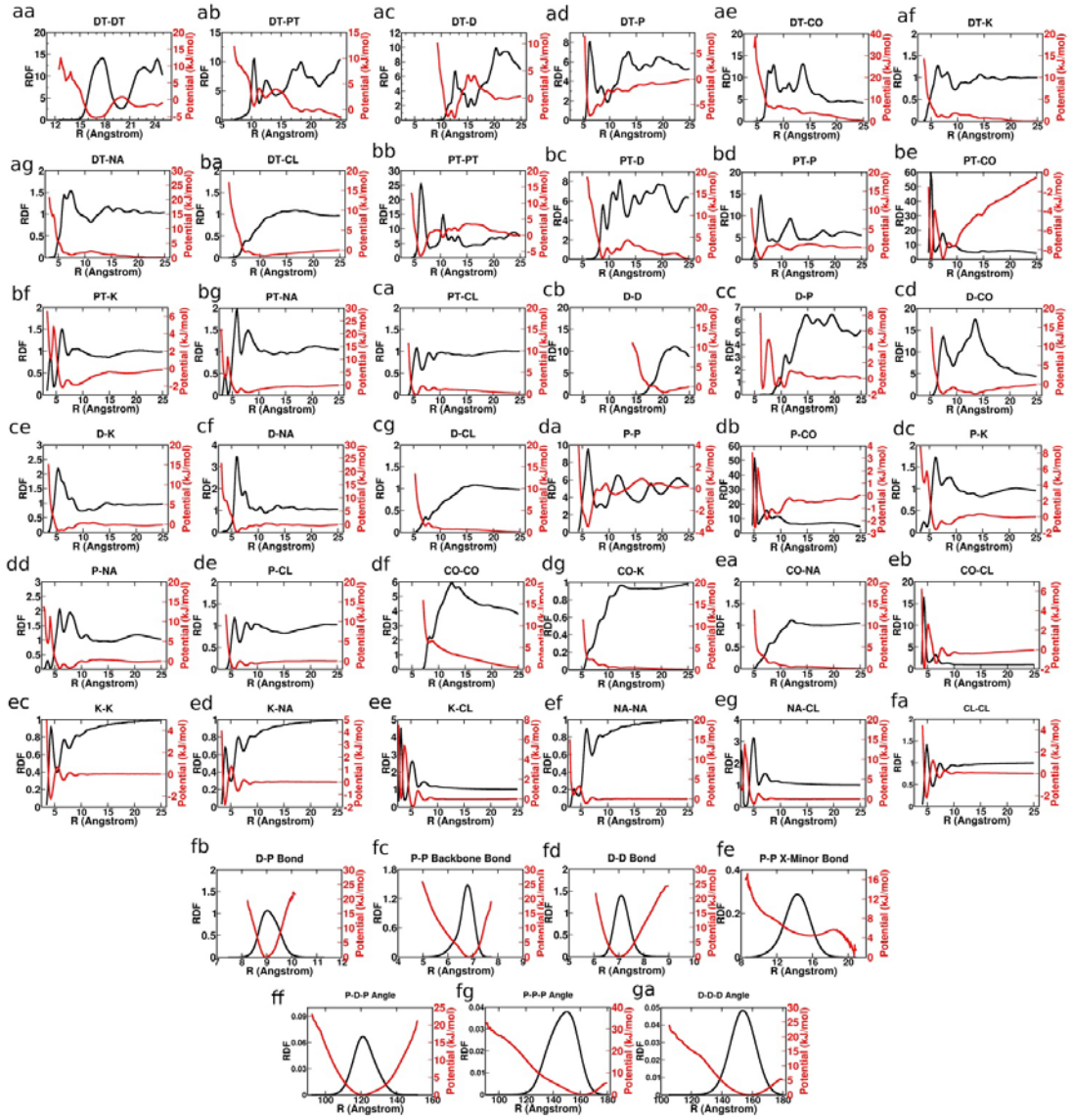

Fig. S2. RDFs (black) and the final effective potentials (red) for all interaction terms in the CG DNA simulations with $CoHex^{3+}$. A total of 36 non-bonded interactions (**aa – fa**), 4 bond interactions (**fb – fe**) and 3 angle interactions (**ff – ga**) are shown.

bioRxiv preprint doi: https://doi.org/10.1101/375626; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

3. Convergence of the Inverse Monte Carlo procedure.

Examples below compare distribution functions and interaction potentials obtained in several iterations of the IMC procedure.

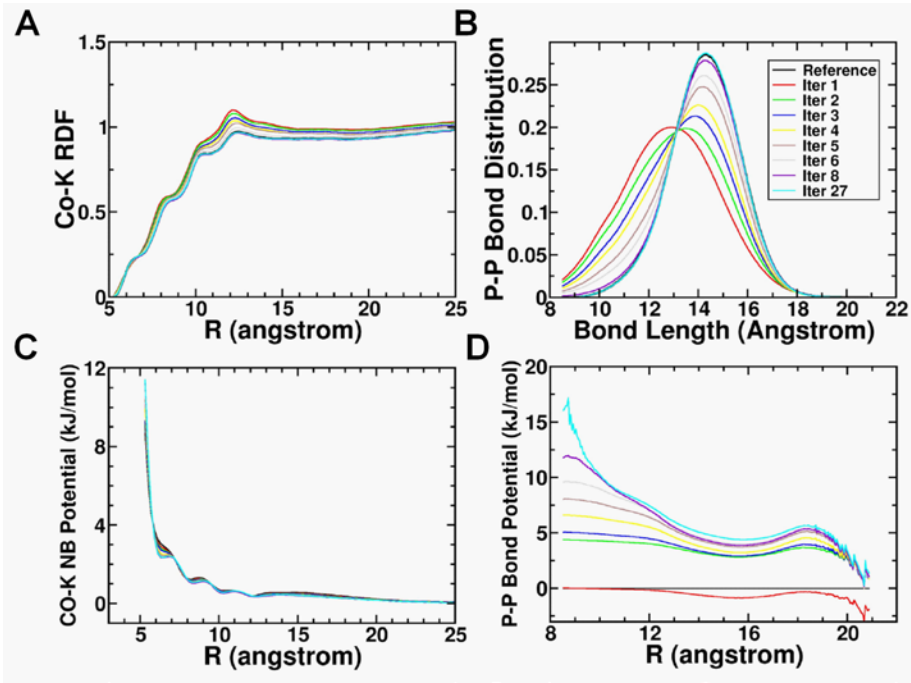

Fig. S3. Selected RDF (**A, B**) and effective potential (**C, D**) plots from several iterations in the IMC calculation. RDFs and potentials of one non-bonded term (Co-K, plots in **A** and **C**) and one bonded term (P-P cross minor groove bond, plots in **B** and **D**) are selected to illustrate 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.

bioRxiv preprint doi: https://doi.org/10.1101/375626; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

4. Transferability of potentials from different systems I.

Fig. S4 demonstrates the transferability of similar interaction types obtained from different underlying systems with different ionic composition and macroscopic properties. It can be noted that the D-Cl potential generated from the DNA system with only NaCl ions in the all-atom setup, displays somewhat different features compared to the same ones generated from the all-atom systems containing either Mg^{2+} or $CoHex^{3+}$. The reason is that in the NaCl-DNA system there is a limited amount of Cl^{-} in the close vicinity of DNA, especially in the groove regions (distances less than 10 Å). In case of Mg^{2+} and $CoHex^{3+}$ counterion systems, some Cl^{-} will be brought to these locations by the presence of these multivalent counterions, which creates a "wave" in the potential at ~7.5 Å.

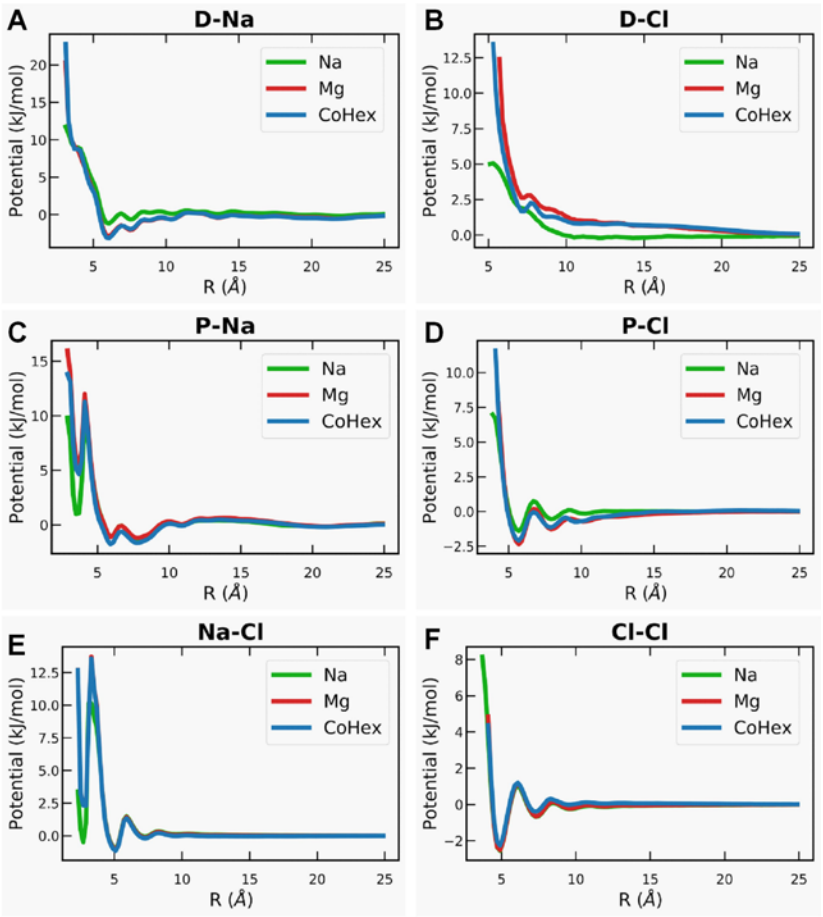

Fig. S4. DNA-ions and ion-ion potentials obtained with the CG DNA model using the IMC procedure applied to three different DNA all-atom MD simulations of different ionic composition. (**A**, **B**) Potentials between DNA central bead (D) and respectively Na^{+} and Cl^{-} ions; (**C**, **D**) potentials between phosphate (P) and respectively Na^{+} and Cl^{-} ions; and (**E**, **F**) potentials between Na^{+} and Cl^{-} ions. Although the individual simulations display highly different behaviour, the potentials are very similar. Blue plots are obtained from the 4DNA + $CoHex^{3+}$ simulation, green with 4DNA + Mg^{2+} simulation and the red with 1DNA + Na^{+} simulation. See Methods for details.

bioRxiv preprint doi: https://doi.org/10.1101/375626; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

5. Transferability of potentials from different systems II.

Fig. S5 demonstrates transferability of the IMC-derived potentials determined from all-atom MD simulations with different ionic conditions and number of DNA molecules in the simulation cell, resulting in very similar behavior in macroscopic behaviour in the CG simulations.

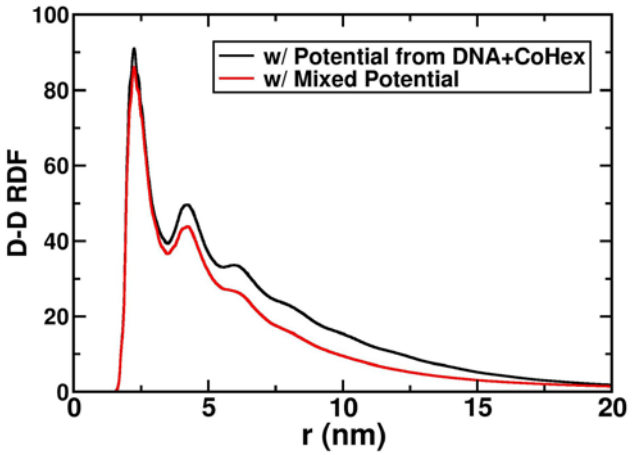

Fig. S5. DNA-DNA distance distribution calculated for the simulations of 25 CG DNA molecules in the presence of $CoHex^{3+}$. Black curve is the data obtained in the simulation using IMC-derived potentials from the all-atom MD simulation with $CoHex^{3+}$ that showed DNA-DNA attraction; red curve is calculated for a system with mixed potentials where potentials for monovalent ions were taken from the Mg^{2+} - DNA all-atom MD simulation with net DNA-DNA repulsion.

bioRxiv preprint doi: https://doi.org/10.1101/375626; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

6. 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. S6B, 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.

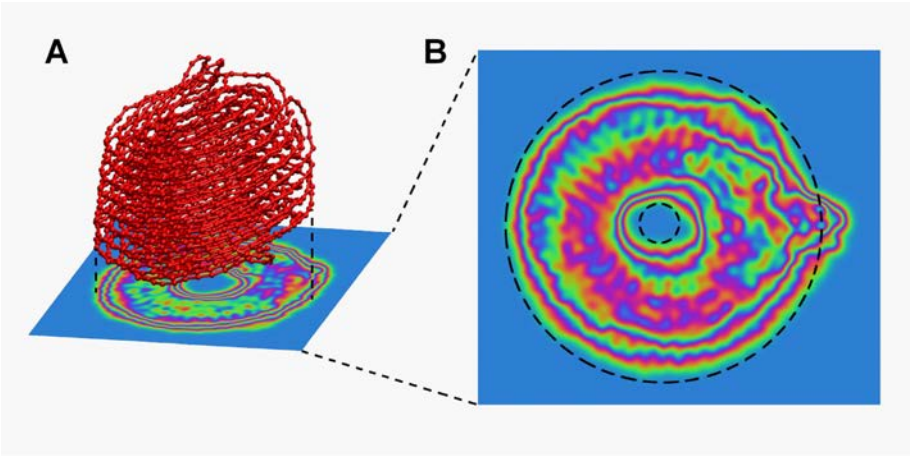

Fig. S6. Measurement of toroid dimensions. First, the toroid is projected to a density map, on a plane perpendicular to the axes through the hole, as shown in (**A**). Then the projected density map was measured using two circles, one to fit the outer perimeter and one to fit the hole (**B**). Finally, 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.

bioRxiv preprint doi: https://doi.org/10.1101/375626; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

7. RDFs and potentials used for the calculation of persistence length

Interaction potentials calculated by the IMC procedure using RDFs obtained in an all-atom MD simulation of one 40 bp DNA in the presence of 130 mM NaCl were used for validation of the CG DNA model (shown in Fig. 1B,D of the main text). These potentials were used to determine salt dependence of the DNA persistence length (see Fig. 2 of the main text).

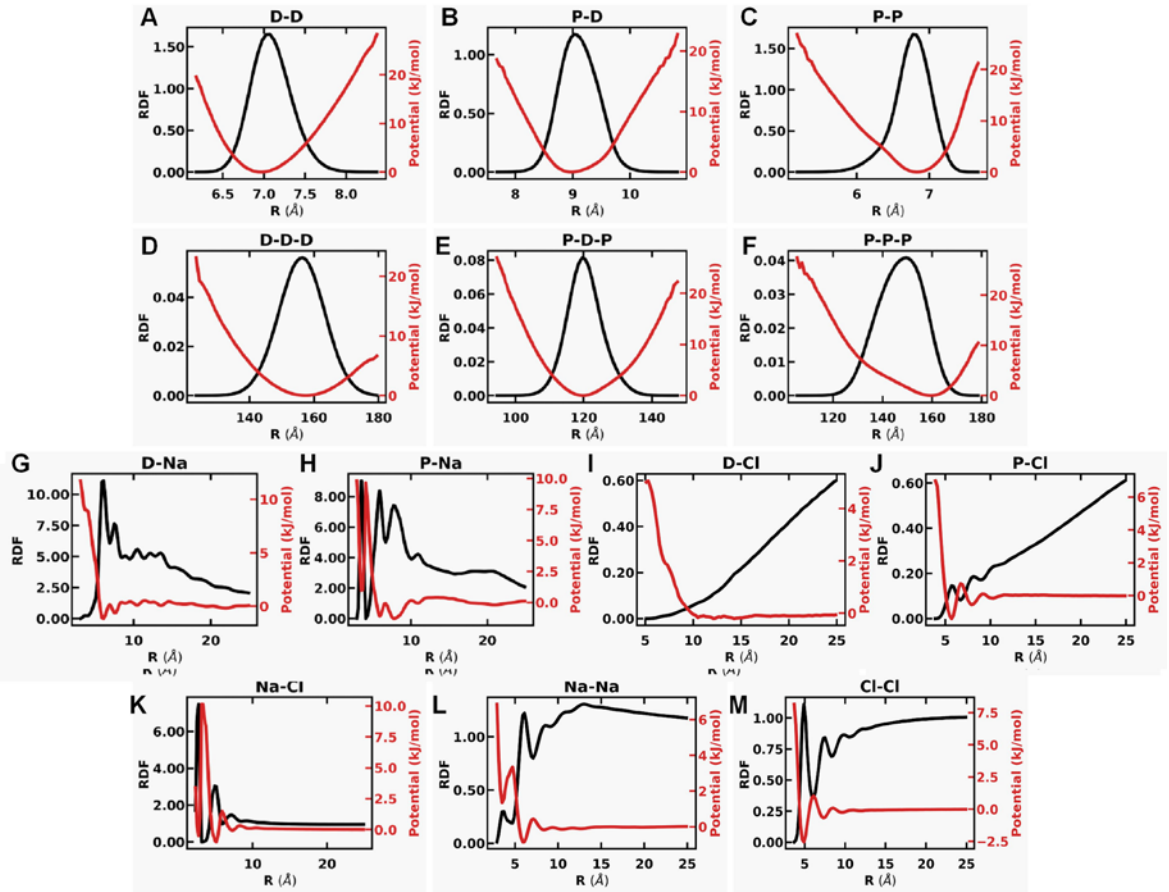

Fig. S7. RDFs (black) and effective potentials (red) for interaction terms in the CG DNA model calculated from all-atom MD simulation of DNA in monovalent salt. Panels **A-C** and **D-F** show data for respectively CG DNA bonds and bonded angles; panels G-J are for the DNA – ion interactions; panels K-M are for the ion – ion RDFs and potentials. Identity of the beads is shown at the top of the graphs.

bioRxiv preprint doi: https://doi.org/10.1101/375626; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Table S1

Table S1. Description of intermediate states during formation of DNA toroid and fiber (see Fig. 6 of the main text).

Snapshot index (Fig. 6A)	Description of structure
a	Extended chain.
b	"Hairpin": there is a turn at the end of DNA, forming a self-contact part consists of two chains of DNA.
c	Three-chain bundle: a bundle of three chains of DNA is formed at the end of DNA.
d	Thick fiber: a bundle of DNA, thicker than **c**, formed at the end of DNA.
e	"Racket": usually on one end of the thick fibre, there's a loop formed by at least two chains of DNA. The shape is similar to a tennis racket without string.
f	Final long and thick fiber: the final fiber we observed, usually longer than d and thicker than **c**.
g	Double-DNA chain loop: a loop consists of double DNA chains, usually formed from b when the loop in b folding back to the double-chain part.
h	Bundle with a side loop: the loop is formed either by newly recruited DNA or "peeling off" of one DNA chain from the bundle.
i	Primer toroid: toroid that has just been formed. It is usually very flexible.
j	Final toroid that is thick and sturdy. Grown from primer toroid by pulling in DNA chain.
k	Single loop: a loop consists of a single DNA chain part and a double DNA chain part.
l	Double loop: two adjacent single loops on DNA chain.