

## RESEARCH ARTICLE

## Ab initio predictions for 3D structure and stability of single- and double-stranded DNAs in ion solutions

Zi-Chun Mu<sup>1,2</sup>✉, Ya-Lan Tan<sup>1</sup>✉, Ben-Gong Zhang<sup>1</sup>, Jie Liu<sup>1</sup>, Ya-Zhou Shi<sup>1</sup>✉\*

**1** Research Center of Nonlinear Science, School of Mathematical & Physical Sciences, Wuhan Textile University, Wuhan, China, **2** School of Computer Science and Artificial Intelligence, Wuhan Textile University, Wuhan, China

✉ These authors contributed equally to this work.

\* [yzshi@wtu.edu.cn](mailto:yzshi@wtu.edu.cn)

## OPEN ACCESS

**Citation:** Mu Z-C, Tan Y-L, Zhang B-G, Liu J, Shi Y-Z (2022) Ab initio predictions for 3D structure and stability of single- and double-stranded DNAs in ion solutions. *PLoS Comput Biol* 18(10): e1010501. <https://doi.org/10.1371/journal.pcbi.1010501>

**Editor:** Alexander MacKerell, University of Maryland School of Pharmacy, UNITED STATES

**Received:** August 22, 2022

**Accepted:** September 27, 2022

**Published:** October 19, 2022

**Copyright:** © 2022 Mu et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant computational data are within the paper and its [Supporting information](#) files. The source code (in C language) of the present model is available at <https://github.com/RNA-folding-lab/DNAfold>. Experimental data are publicly available from published papers cited in the work. The 3dRNA/DNA web server used in this work is available by <http://biophy.hust.edu.cn/new/3dRNA/create>.

**Funding:** This work was supported by the Grants from the National Science Foundation of China (11971367 to BGZ, and 11605125 to YZS). The

## Abstract

The three-dimensional (3D) structure and stability of DNA are essential to understand/control their biological functions and aid the development of novel materials. In this work, we present a coarse-grained (CG) model for DNA based on the RNA CG model proposed by us, to predict 3D structures and stability for both dsDNA and ssDNA from the sequence. Combined with a Monte Carlo simulated annealing algorithm and CG force fields involving the sequence-dependent base-pairing/stacking interactions and an implicit electrostatic potential, the present model successfully folds 20 dsDNAs ( $\leq 52$ nt) and 20 ssDNAs ( $\leq 74$ nt) into the corresponding native-like structures just from their sequences, with an overall mean RMSD of 3.4Å from the experimental structures. For DNAs with various lengths and sequences, the present model can make reliable predictions on stability, e.g., for 27 dsDNAs with/without bulge/internal loops and 24 ssDNAs including pseudoknot, the mean deviation of predicted melting temperatures from the corresponding experimental data is only  $\sim 2.0^\circ\text{C}$ . Furthermore, the model also quantitatively predicts the effects of monovalent or divalent ions on the structure stability of ssDNAs/dsDNAs.

## Author summary

To determine 3D structures and quantify stability of single- (ss) and double-stranded (ds) DNAs is essential to unveil the mechanisms of their functions and to further guide the production and development of novel materials. Although many DNA models have been proposed to reproduce the basic structural, mechanical, or thermodynamic properties of dsDNAs based on the secondary structure information or preset constraints, there are very few models can be used to investigate the ssDNA folding or dsDNA assembly from the sequence. Furthermore, due to the polyanionic nature of DNAs, metal ions (e.g.,  $\text{Na}^+$  and  $\text{Mg}^{2+}$ ) in solutions can play an essential role in DNA folding and dynamics. Nevertheless, *ab initio* predictions for DNA folding in ion solutions are still an unresolved problem. In this work, we developed a novel coarse-grained model to predict 3D structures and

fundamental role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

thermodynamic stabilities for both ssDNAs and dsDNAs in monovalent/divalent ion solutions from their sequences. As compared with the extensive experimental data and available existing models, we showed that the present model can successfully fold simple DNAs into their native-like structures, and can also accurately reproduce the effects of sequence and monovalent/divalent ions on structure stability for ssDNAs including pseudoknot and dsDNAs with/without bulge/internal loops.

## Introduction

DNA can adopt many structures beyond the right-handed B-form double-helices, which takes it far beyond being the molecule that stores and transmits genetic information in biological systems [1,2]. Some non-B-form DNAs within the human genes, such as hairpins, triplexes, Z-DNA, quadruplexes, and i-motifs, have been proposed to participate in several biologically important processes (e.g., regulation and evolution), leading to mutations, chromosomal translocations, deletions and amplifications in cancer and other human diseases [1–4]. Furthermore, self-assembled functional DNA structures have proven to be excellent materials for designing and implementing a variety of nanoscale structures and devices, including interlocked, walkers, tweezers, shuttles, logic circuits, and origami, which have promising applications ranging from photonic devices to drug delivery [5–8]. Since short double- and single-stranded DNA (dsDNA and ssDNA) structures (e.g., duplex, hairpins, pseudoknots, and junctions) are essential to build blocks for the construction of non-B-form DNAs and various nano-architectures, advancement in the knowledge of structures and key properties (e.g., thermodynamics and mechanics) for these DNAs will be helpful to understand and ultimately control their biological functions and guide the production and development of novel materials [7–9].

Although several experimental methods such as cryo-electron microscopy, X-ray crystallography, NMR spectroscopy, and other single-molecule techniques (e.g., optical/magnetic tweezers and atomic force microscopy) can be used to determine three-dimensional (3D) structures or elastic properties of DNAs [10–15], there are still full of challenges (e.g., time-consuming and expensive) to experimentally provide insight into DNA folding/hybridization. Thus, the field of computer simulation is rapidly evolving to provide finer details on key features of DNA biophysics compared to experimental approaches [16–19]. For example, all-atom molecular dynamics (MD) simulations based on force fields, such as CHARMM and AMBER, have been widely used to investigate dynamics, flexibility, mechanics, or form transition of dsDNA helices at the microscopic level [20–24]. However, due to the innumerable degrees of freedom, the MD simulations are limited to small DNAs and to short times even with an advanced-sampling approach and parallel tempering scheme [16,24–26].

On the other hand, the simple continuum DNA models (e.g., worm-like chain model), which treat the double helix as a continuous elastic rod with bending and torsional stiffness, are commonly used to well describe mechanical behavior or elastic bending of dsDNA on long length-scales [27–30]. Correspondingly, the nearest-neighbor model can predict secondary structures and melting profiles (e.g., free energy and melting temperature) for ssDNA and dsDNA through the combination of free energy minimization, partition function calculations, and stochastic sampling [9,31]. However, these simple models are unable to provide any 3D structure information on DNAs.

Therefore, many coarse-grained (CG) DNA models, which represent DNA using a reduced number of interaction sites while striving to keep the important details, have been developed

in recent years to model 3D structures or thermodynamic and structural properties of DNAs [32–39]. For example, by mapping each nucleotide into six to seven CG beads, the Martini model combined with MD simulations opens the way to perform large-scale modeling of complex biomolecular systems involving DNA, such as DNA–DNA and DNA–protein interactions [40,41]. Very recently, a three-bead CG model, MADna, was developed to reproduce the conformational and elastic features of dsDNA, including the persistence length, stretching/torsion moduli, and twist–stretch couplings [42]. However, since the two models need constraints (e.g., predefined elastic network or bonded interactions for paired bases) to maintain a double helix, they cannot be used to study DNA hybridization, melting, and hairpin formation [40–42].

Moreover, some other Go-like models, including 3SPN, oxDNA, and TIS, have been proposed to fill the gap [43–50]. The 3SPN model, which reduces the complexity of a nucleotide to three interactions sites (i.e., phosphate, sugar, and base), can successfully capture DNA denaturation/renaturation and provide a reasonable description of other thermomechanical and structural properties for DNAs (e.g., persistence length, bubble formation, major and minor groove widths, and local curvature) by involving in base-stacking and base-pairing interactions [43–45]. The oxDNA model uses three collinear sites and a vector normal to the base site to construct the angle-dependent potentials including coplanar base-stacking and linear hydrogen bonding interactions, which are parametrized to accurately describe basic structural, mechanical, and thermodynamic properties of ss/dsDNA [46–48]. More significantly, with fine-tuned structural parameters, the model can also treat large DNA nanostructures, such as DNA origami and nanotweezers [48,49]. The TIS-DNA is another robust three-interaction-site CG model, and using a set of nucleotide-specific stacking parameters obtained from thermodynamic properties of dimers, the model can reproduce the sequence-dependent mechanical, as well as thermodynamic properties of DNAs, covering the force-extension behavior and persistence lengths of poly(dA)/poly(dT) chains, elasticity of dsDNA and melting temperatures of hairpins [50–52]. The use of Go-like interactions (e.g., non-bonded potentials to penalize deviations from a reference structure) can effectively restrict the range of conformations that may be sampled by the CG model, and simultaneously, it also limits the possibility of the model on structure prediction from the sequence.

Recently, the 3dRNA/DNA web server was further developed based on the 3dRNA to build three-dimensional (3D) structures of RNA and DNA from template segments with very high accuracy using sequence and secondary structure information [53–55]. Similarly, a pipeline presented by Jeddi and Saiz can also be used to predict DNA hairpins by integrating the existed 2D and 3D structural tools (e.g., Mfold, Assemble, and MD) [56]. However, the two structure prediction methods are dependent on secondary structures, while there is still a problem to exactly predict secondary structures of DNAs [31]. Fortunately, a minimal physics-based CG model of nucleic acids named NARES-2P was proposed to fold dsDNA from separate strands without any Go-like potentials and secondary structure information. Although the model was constructed using the bottom-up strategy, where each component of the energy function was fitted separately to the respective potential of mean force obtained from all-atom potential-energy surfaces, it can reproduce many properties of double-helix B-DNA, such as duplex formation, melting temperatures, and mechanical stability [57–59]. Contrary to the oversimplification of NARES-2P (i.e., two sites per nucleotide), the HiRE-RNA is an empirical CG model for RNA and DNA, whose resolution is high enough (i.e., six or seven beads for each nucleotide) to preserve many important geometric details, such as base pairing and stacking. Without imposing preset pairings for the nucleotides, the HiRE-RNA can investigate both dynamical and thermodynamic aspects of dsDNA assemblies, as well as the effect of sequences on the melting curves of the duplexes [60,61]. Despite the advances, the parameters of the two models

may need further validation for quantifying thermodynamic and 3D structure to accord with experiments, especially for ssDNA.

In addition, due to the polyanionic nature of DNAs, metal ions (e.g., Na<sup>+</sup> and Mg<sup>2+</sup>) in solutions can play an essential role in DNA folding and dynamics [12–14,62–65]. Although several of the existing models such as 3SPN, oxDNA, TIS, and NARES-2P have taken the electrostatic interactions into account using the Debye-Huckel approximation or mean-field multipole–multipole potentials and reproduced some monovalent salt-dependent structural properties (e.g., persistence length, torsional stiffness or melting temperatures) of DNAs [45,48,50,58], quantitatively predicting the 3D structure and thermodynamic stability for DNA including both ssDNA and dsDNA in ion solutions (especially divalent ions) from the sequence is still an unresolved problem. Recently, we have proposed a three-bead CG model to simulate RNA folding from the sequence, and with an implicit electrostatic potential, the model can make reliable predictions on 3D structures and stability for RNA hairpins, pseudoknots, and kissing complexes in ion solutions [66–70]. However, due to the differences in geometry, base stacking strength, and flexibility between DNA and RNA, the present model cannot be directly used to simulate DNA folding.

In this work, we further developed an *ab initio* CG model of DNA to predict the 3D structure, stability, and salt effect for both dsDNA and ssDNA. First, the bonded and nonbonded potentials were parameterized based on the statistical analysis of known DNA 3D structures, as well as experimental thermodynamic parameters and melting data. Afterward, the model was validated through 3D structure and stability predictions for DNAs including double helices, hairpins, and pseudoknots with different lengths and sequences, as compared with the extensive experimental data. Furthermore, we also showed that the effects of monovalent and divalent ions on DNA structure stability predicted by the present model are in accordance with the corresponding experiments.

## Materials and methods

### CG structure representation for DNAs

To be consistent with our previous RNA CG model [66], each nucleotide in DNA is also simplified into three beads: P, C, and N, to represent the phosphate group, sugar ring, and base plane, respectively. For simplicity, the three beads are placed at the positions of existing atoms (i.e., P, C4', and N1 for pyrimidine or N9 for purine) (Fig 1) and are treated as van der Waals spheres with the radii of 1.9 Å, 1.7 Å and 2.2 Å, respectively [66,70]. One unit negative charge ( $-e$ ) is placed on the center of P bead to describe the highly charged nature of DNA.

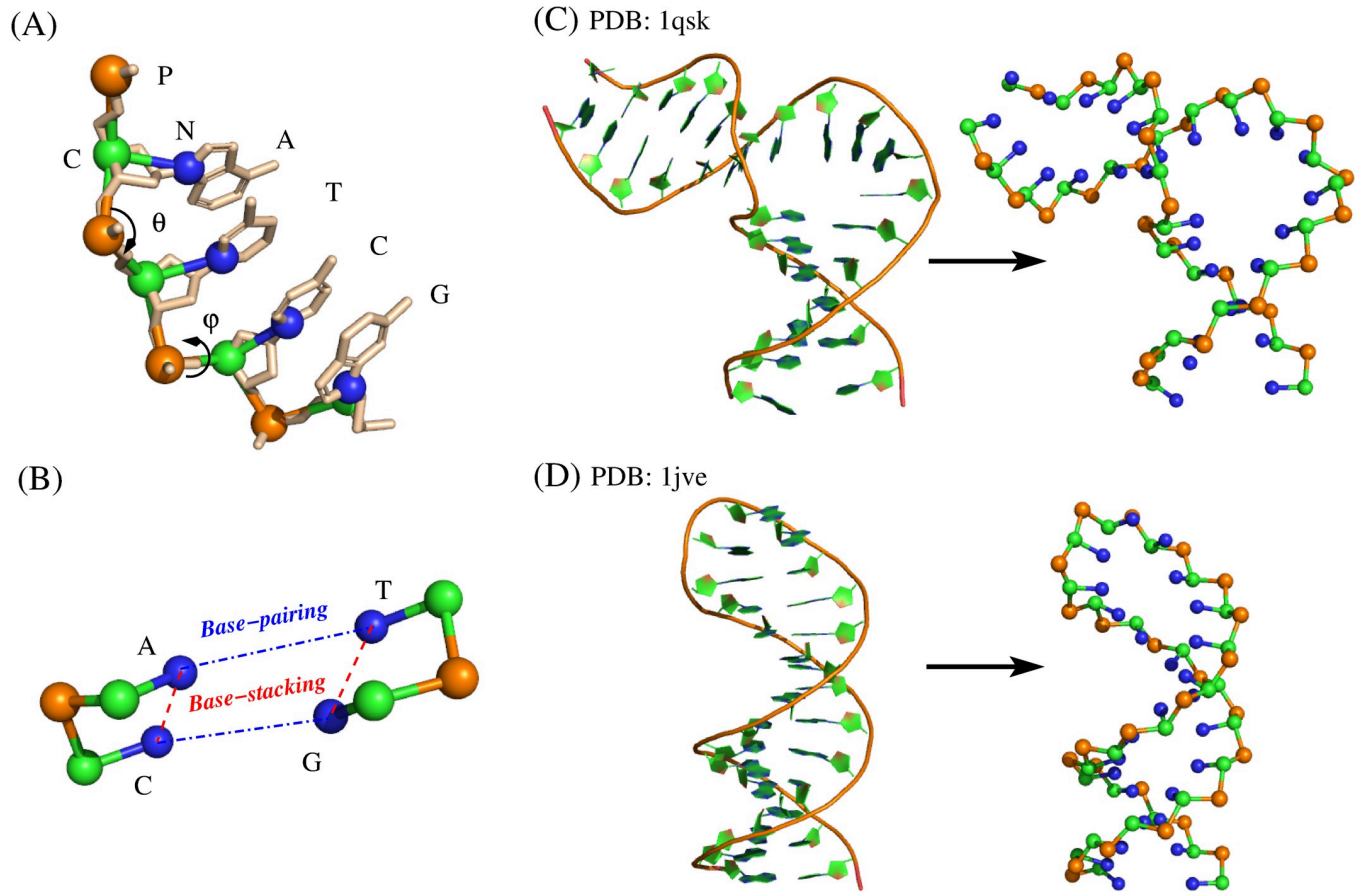
### Energy functions

The total energy  $U$  in the present DNA CG model is composed of the following eight components:

$$U = U_b + U_a + U_d + U_{exc} + U_{bp} + U_{bs} + U_{cs} + U_{el}. \quad (1)$$

The first three terms are bonded potentials for virtual bonds  $U_b$ , bond angles  $U_a$ , and dihedrals  $U_d$ , respectively, which are used to mimic the connectivity and local geometry of DNA chains. The function forms of these terms are listed in S1 Text, which can also be found elsewhere [44,46,50,66].

The remaining terms of Eq 1 describe various pairwise, nonbonded interactions. The  $U_{exc}$  represents the excluded volume interaction between the CG beads and it is modeled by a purely repulsive Lennard-Jones potential. The  $U_{bp}$  in Eq 1 is an orientation-dependent base-



**Fig 1. The representations of all-atom and CG model for each deoxynucleotide and ss/dsDNA molecules, as well as the schematics of base-pairing and base-stacking in the present model.** (A) Our coarse-grained representation of a DNA fragments including deoxynucleotides of A, T, G, and C superimposed on the all-atom representation. The three beads are located at phosphate (P), sugar (C4') and pyrimidine (N1) or purine (N9).  $\theta$  and  $\phi$  are the schematics of CG bonded angle (CPC) and dihedral (CPCP), respectively. (B) The schematic representation of base-pairing (blue) and base-stacking (red) interaction. (C,D) The 3D structures of (C) a dsDNA with bulge loop (PDB:1qsk) and (D) an ssDNA hairpin (PDB: 1jve) in all-atomistic (left) and our CG representation (right). The 3D structures are shown with PyMol (<http://www.pymol.org>).

<https://doi.org/10.1371/journal.pcbi.1010501.g001>

pairing interaction for the possible canonical Watson-Crick base pairs (G-C and A-T). The formula of  $U_{bp}$  is similar to the form of hydrogen-bonding interaction used in the TIS model [50], and the backbone dihedrals are replaced by two simpler distances between CG beads in pairing nucleotides to describe the orientation of hydrogen-bonding interactions; see Eq S6 in S1 Text. The sequence-dependence base-stacking interaction  $U_{bs}$  between two nearest neighbour base pairs is given by

$$U_{bs} = \frac{1}{2} \sum_{ij}^{N_{st}} |G_{i,i+1,j-1,j}| \left\{ \left[ 5 \left( \frac{\sigma_{st}}{r_{i,i+1}} \right)^{12} - 6 \left( \frac{\sigma_{st}}{r_{i,i+1}} \right)^{10} \right] + \left[ 5 \left( \frac{\sigma_{st}}{r_{j,j-1}} \right)^{12} - 6 \left( \frac{\sigma_{st}}{r_{j,j-1}} \right)^{10} \right] \right\} \quad (2)$$

where  $\sigma_{st}$  is the optimum distance of two neighbour bases in the known DNA helix structures.  $G_{i,i+1,j-1,j}$  in Eq 2 is the strength of base-stacking energy, and it can be calculated by  $G_{i,i+1,j-1,j} = \Delta H - T(\Delta S - \Delta S_c)$ . Here,  $T$  is the absolute temperature in Kelvin,  $\Delta H$  and  $\Delta S$  are the DNA thermodynamic parameters derived from experiments [9,71], and  $\Delta S_c$  is the conformational entropy change that is naturally included in the Monte Carlo (MC) simulations, during the formation of one base pair stacking; see more details in Eq S7 in S1 Text as well as the previous

works [66,70,72,73]. In addition, the coaxial-stacking interaction  $U_{cs}$  between two discontinuous neighbour helices is also taken into account by the present model through calculating the stacking potential of the interfaced base-pairs, and the expression can be found in Eq S10 in S1 Text.

The last term  $U_{el}$  in Eq 1 is used to calculate the electrostatic interactions between phosphates groups (e.g.,  $i$ -th and  $j$ -th P beads), and it is given by

$$U_{el} = \sum_{i < j}^{N_p} \frac{(Qe)^2}{4\pi\epsilon_0\epsilon(T)r_{ij}} e^{-\frac{r_{ij}}{l_D}} \quad (3)$$

where  $e$  is the elementary charge,  $r_{ij}$  is the distance between  $i$ - and  $j$ -th P beads, and  $N_p$  is the total number of P beads in a DNA.  $l_D$  is Debye length, which defines the ionic screening at different solution ionic strengths.  $\epsilon_0$  and  $\epsilon(T)$  are the permittivity of vacuum and an effective temperature-dependent dielectric constant, respectively [50,66,67].  $Q$  is the reduced charge fraction derived based on Manning's counterion condensation theory and the tightly bound ion model [74–76]; see Eq S11 in S1 Text. Due to the inclusion of the  $U_{eb}$ , the present model can be used to study DNA folding in pure (e.g.,  $\text{Na}^+$ ) as well as mixed (e.g.,  $\text{Na}^+/\text{Mg}^{2+}$ ) ion solutions.

## Parametrization

The initial parameters of bonded potentials (i.e.,  $U_b$ ,  $U_a$ , and  $U_d$  in Eq 1) were derived from the Boltzmann inversion of the corresponding CG atomistic distribution functions, obtained by the statistical analysis on experimental DNA structures in the Protein Data Bank (PDB) (<http://www.rcsb.org/pdb/-home/home.do>) (Fig A in S1 Text). First, 752 pure DNA structures (10nt–200nt) with resolution  $<3.5\text{\AA}$  were collected from the PDB, and then, the DNAs with triplex, quadruple helices or unnatural structures were further removed. In addition, excluding the DNA structures with sequence identity  $>80\%$  and the ssDNAs/dsDNAs used for model validation on 3D structure prediction, there were only 138 DNA structures can be used to parameterize the energy potentials, and the PDB codes of these DNAs are listed in Table A in S1 Text. Since the known DNA structures are generally double helices, the initial parameters from these structures could not be reasonable to describe DNA chains during folding processes. In our previous RNA model, two sets of parameters ( $\text{Para}_{\text{helical}}$  and  $\text{Para}_{\text{nonhelical}}$ ) were calculated from stems and loops in experimental structures, respectively [66,67], and the  $\text{Para}_{\text{nonhelical}}$  ones were used to successfully describe the folding of an RNA from a free chain. However, due to the limitation of the number of loop regions in known DNA structures, obtaining suitable parameters for DNA-free chains directly from these structures is unrealistic. Although we also did MD simulations for unstructured ssDNA and tried to extract the bonded parameters from the conformations (Fig B in S1 Text), because of some differences in optimum values of several angles between experimental and MD simulated structures, we gave them up. Eventually, based on the distributions of bond length/angle/dihedral for nonhelical parts in RNA structures are just slightly broader than that of helical parts [66], we simply set the strengths of DNA bonded potentials in  $\text{Para}_{\text{nonhelical}}$  as one-half of that in  $\text{Para}_{\text{helical}}$ . Note that, the  $\text{Para}_{\text{nonhelical}}$  is used in the folding process, and the  $\text{Para}_{\text{helical}}$  is only used for stems during folded structure refinement. Whereafter, the initial parameters were further optimized through the comparisons between the simulated and experimental bond length/angle distributions [34,77], and in this process, there are only two dsDNAs (PDB codes: 1agh, 3bse) and two ssDNAs (PDB codes: 1ac7, 1jve) were used.

For nonbonded potentials, the geometric parameters in base-pairing/stacking functions were obtained from the known structures; see Fig C in S1 Text. The strength of base-stacking

**Table 1.** The melting temperatures ( $T_m$ ) for dsDNAs at 1M [Na<sup>+</sup>].

dsDNAs	Sequence <sup>a</sup>	Refs.	Length (nt)	$C_s$ (mM)	Expt. (°C)	Pred. (°C)	Deviation (°C)
DS1	(GCGC) <sub>2</sub>	[71]	8	0.1	27.5	26.6	-0.9
DS2	(CGCG) <sub>2</sub>	[71]	8	0.1	23.7	22.5	-1.2
DS3	(CGCGCG) <sub>2</sub>	[71]	12	0.1	55.7	54.0	-1.7
DS4	(GACGTC) <sub>2</sub>	[71]	12	0.1	36.1	36.3	0.2
DS5	(GGCGCC) <sub>2</sub>	[71]	12	0.1	53.9	53.0	-0.9
DS6	(CCGCGG) <sub>2</sub>	[71]	12	0.1	55.2	53.6	-1.6
DS7	(GCTAGC) <sub>2</sub>	[71]	12	0.1	34.3	36.6	2.3
DS8	(CGATATCG) <sub>2</sub>	[71]	16	0.1	44.1	47.0	2.9
DS9	(GGACGTCC) <sub>2</sub>	[71]	16	0.1	55.1	56.0	0.9
DS10	(CATATGGCCATATG) <sub>2</sub>	[71]	28	0.1	65.3	62.8	-2.5
DS11	GGGACC/GGTCCC	[71]	12	0.4	44.9	45.0	0.1
DS12	CAAAAAG/CTTTTGG	[71]	14	0.4	31.5	33.0	1.5
DS13	CAAAAAG/CTTTTTG	[71]	16	0.4	36.9	39.0	2.1
DS14	CAAACAAAG/CTTTGTTG	[71]	18	0.4	47.3	49.1	1.8
DS15	CCATCGCTACC/GGTAGCGATGG	[71]	22	0.4	67.6	64.6	-3.0
DS16	GCGAAAAGCG/CGCTTTTCGC	[71]	20	0.4	65.2	66.4	1.2
DS17	(GCATGC) <sub>2</sub>	[86]	12	0.1	38.3	38.0	-0.3
DS18	GCGAATAAGCG/CGCTTTTCGC	[87]	21	0.1	46.3	46.6	0.3
DS19	GCGAAAAGCG/CGCTTCTTCGC	[87]	21	0.1	44.6	46.4	1.8
DS22	(CGATTATCG) <sub>2</sub>	[88]	18	0.1	32.4	34.8	2.4
DS21	(GGACTGTCC) <sub>2</sub>	[88]	18	0.1	44.5	46.2	1.7
DS22	CTCGTCCAGTGC/GCACTGGACGAG	[89]	24	0.01	66.0	65.2	-0.8
DS23	CTCGTCCAGTGC/GCACTG(T) <sub>1</sub> GACGAG	[89]	25	0.01	59.2	62.0	2.8
DS24	CTCGTCCAGTGC/GCACTG(T) <sub>6</sub> GACGAG	[89]	30	0.01	46.2	48.0	1.8
DS25	CTCGTC(T) <sub>1</sub> CAGTGC/GCACTG(T) <sub>1</sub> GACGAG	[89]	26	0.01	59.7	57.9	-1.8
DS26	CTCGTC(T) <sub>2</sub> CAGTGC/GCACTG(T) <sub>2</sub> GACGAG	[89]	28	0.01	54.9	56.6	1.7
DS27	CTCGTC(T) <sub>6</sub> CAGTGC/GCACTG(T) <sub>6</sub> GACGAG	[89]	36	0.01	46.3	44.0	-2.3

<sup>a</sup> The subscript of “2” represents the self-complementary sequences, and for non-self-complementary sequences, the two strands are separated by ‘/’ and both start from 5’. The loop/unpaired nucleotides are underlined, and the “(T)<sub>N</sub>” means that the nucleotide T repeats N times.

<https://doi.org/10.1371/journal.pcbi.1010501.t001>

was estimated from the combination of the experimental thermodynamics parameters and MC simulations; see Eqs S7-S9 and Fig D in [S1 Text](#). The strength of base-pairing (i.e.,  $\epsilon_{bp}$  in Eq S6 in [S1 Text](#)) was determined by comparing the predicted melting temperatures ( $T_m$ 's) of four ss-/dsDNAs with corresponding experiments. That is, for two ssDNA hairpins (sequences: GCGCTTTTTGCGC and GGAGCTTTTTGCTGC; ion condition: 1M NaCl; see [Table 1](#)) and two dsDNAs (sequences: GCTAGC/GCTAGC and GGGACC/GGTCCC; strand concentration: 0.1mM and 0.4mM, respectively; ion condition: 1M NaCl; see [Table 2](#)), we used the present model to predict their  $T_m$ 's through continuously adjusting the  $\epsilon_{bp}$  until the agreement between predicted and experimental data is satisfactory.

The detailed descriptions, as well as the parameters of all the potentials in [Eq 1](#), can be found in [S1 Text](#).

## Simulation procedure

During DNA folding from sequence without any preset constraints, it is easy to fall into a metastable state with local minimum energy. To effectively avoid that, the MC simulated

Table 2. The melting temperatures ( $T_m$ ) for single-stranded DNAs at given ion conditions.

ssDNAs	Sequence <sup>a</sup>	Refs.	Length (nt)	[Na <sup>+</sup> ] (M)	Expt. (°C)	Pred. (°C)	Deviation (°C)
SS1	GCGC(T) <sub>5</sub> GCGC	[91]	13	1.0	76.4	78.4	2.0
SS2	GCCGC(T) <sub>5</sub> GCCGC	[91]	15	1.0	81.6	84.5	2.9
SS3	GCAGC(T) <sub>5</sub> GCTGC	[91]	15	1.0	74.9	75.8	0.9
SS4	GCTGC(T) <sub>5</sub> GCGC	[91]	15	1.0	58.0	59.0	1.0
SS5	GCTGC(T) <sub>5</sub> GCTGC	[91]	15	1.0	56.9	55.1	-1.8
SS6	GCGC(T) <sub>3</sub> GCGC	[90]	11	0.1	80.7	78.0	-2.7
SS7	GCGC(T) <sub>5</sub> GCGC	[90]	13	0.1	73.3	71.9	-1.4
SS8	GCGC(T) <sub>7</sub> GCGC	[90]	15	0.1	62.8	63.9	1.1
SS9	GCGC(T) <sub>9</sub> GCGC	[90]	17	0.1	56.2	58.1	1.9
SS10	GTAC(T) <sub>5</sub> GTAC	[89]	13	0.1	33.4	32.2	-1.2
SS11	ATCCTA(T) <sub>5</sub> TAGGAT	[90]	17	0.2	51.0	54.0	3.0
SS12	ATCCTA(T) <sub>6</sub> TAGGAT	[90]	18	0.2	48.1	51.6	3.5
SS13	ATCCTA(T) <sub>7</sub> TAGGAT	[90]	19	0.2	43.9	47.1	3.2
SS14	GAATTC(T) <sub>5</sub> GAATTC	[94]	17	0.2	55.3	53.1	-2.2
SS15	CCCAA(T) <sub>12</sub> TTGGG	[95]	22	0.25	57.9	59.2	1.1
SS16	CCCAA(T) <sub>16</sub> TTGGG	[95]	26	0.25	51.7	55.5	3.8
SS17	CGGATAA(T) <sub>4</sub> TTATCCG	[96]	18	0.1	63.4	66.0	2.6
SS18	CGGATAA(T) <sub>8</sub> TTATCCG	[96]	22	0.1	50.5	52.5	2.0
SS19	CGGATAA(T) <sub>12</sub> TTATCCG	[96]	26	0.1	41.1	42.0	0.9
SS20	CGGATAA(T) <sub>16</sub> TTATCCG	[96]	30	0.1	35.1	34.1	-1.0
SS21	CGGATAA(T) <sub>20</sub> TTATCCG	[96]	34	0.1	29.6	26.8	-2.8
SS22	AAAAAA(C) <sub>5</sub> TTTTTTT	[36]	19	1.0	55.0	58.5	3.5
SS23	CGCGCG(T) <sub>4</sub> GAAATTCGCGCG(T) <sub>6</sub> AATTC	[90]	32	0.1	52.6/ 70.7	48.8/72.0	-3.8/ 1.3
SS24	GCGC(T) <sub>5</sub> GCGCGTAC(T) <sub>5</sub> GTAC	[92]	26	0.1	50.1/ 76.1	46.9/75.0	-2.1/ -1.1

<sup>a</sup> The sequences start from 5', the loop nucleotides are underlined, and the "(T)<sub>N</sub>" means that the nucleotide T repeats N times.

<https://doi.org/10.1371/journal.pcbi.1010501.t002>

annealing algorithm, whose capacity has been proved in protein/RNA folding, was used to sample conformations for ssDNA or dsDNA [66,78,79]. For each DNA, a random chain configuration is generated from its sequence, and for dsDNA, the two chains are separately placed in a cubic box, the size of which is determined by the concentration of a single strand. Afterward, the simulation of a DNA system with a given monovalent/divalent ion condition is performed from a high temperature (e.g., 120°C) to the target temperature (e.g., room/body temperature). At each temperature, conformational changes are accomplished via the translation and pivot moves, which have been demonstrated to be rather efficient in sampling conformations of polymers [80,81], and the changes are accepted or rejected according to the standard Metropolis algorithm [66,70]. The equilibrium conformations at different temperatures during the cooling process are used to analyze the stability of the DNA. In structure prediction, the last conformation at the target temperature is taken as the initial predicted structure, which can be further refined to better capture the geometry of helical parts by introducing the bonded parameters of Para<sub>helical</sub> for consecutive base-pairing regions. After structure refinement, an ensemble of structures would be obtained, and the mean RMSD (the averaged value over the whole structure ensemble) and minimal RMSD (corresponding to the structure closest to the native one) calculated over CG beads from the corresponding atoms in the native PDB structure is used to evaluate the reliability of the present model on DNA 3D structure prediction.



## Calculation of melting temperature

At each temperature, the fractions of folded (F, consistent secondary structures with predicted at lowest temperature) and unfolded (U, no more than one base pair) states could be fitted to a two-state model through the following equations [9,66]:

$$f_F(T) = \frac{1}{1 + e^{(T-T_{m1})/dT_1}} \quad (4)$$

$$f_U(T) = \frac{1}{1 + e^{(T-T_{m2})/dT_2}} \quad (5)$$

where  $T_{m1}$  and  $T_{m2}$  are the two melting temperatures of the corresponding transitions (folded state to possible intermediate state (I) and intermediate state to unfolded state), respectively.  $dT_1$  and  $dT_2$  are the corresponding adjustable parameters. Based on the  $f_F(T)$  and  $f_U(T)$ , the fraction of the number of denatured base pairs  $f(T)$  can be calculated by [69]

$$f(T) = 1 - [(1 - f_I) \cdot f_F(T) + f_I \cdot (1 - f_U(T))] \quad (6)$$

Here,  $f_I$  is the fraction of the number of denatured base pairs when the fraction for the I state is maximum. And then, the  $df/dT$  (the first derivative of  $f$  with respect to temperature) profile can be calculated to compare with the corresponding experimental data. It should be noted that for simple hairpins and short duplexes used in this work, the I state almost never occurs, and the  $f_I$  in Eq 6 could be set to 0, which means that  $f_U(T)$  is approximately equal to  $1 - f_F(T)$  and only one  $T_m$  can be obtained.

To improve the simulation efficiency for dsDNA with low strand concentrations  $c_s$  (e.g.,  $<0.1$  mM), the MC simulations were performed at a relatively high strand concentration  $c_s^h$  (e.g., 1 mM), and the fraction ( $f(T; c_s)$ ) of denatured base pairs at lower  $c_s$  can be calculated by that ( $f(T; c_s^h)$ ) at  $c_s^h$  [70,82,83]

$$f(T; c_s) = \frac{\lambda f(T; c_s^h)}{1 + (\lambda - 1)f(T; c_s^h)} \quad (7)$$

where  $\lambda = c_s^h/c_s$ . Furthermore, for a dsDNA with a two-state transition, the melting temperature  $T_m(c_s)$  at  $c_s$  can be directly obtained from  $T_m(c_s^h)$  based on Eqs 4–7,

$$T_m(c_s) = T_m(c_s^h) - dT_1 \ln \lambda \quad (8)$$

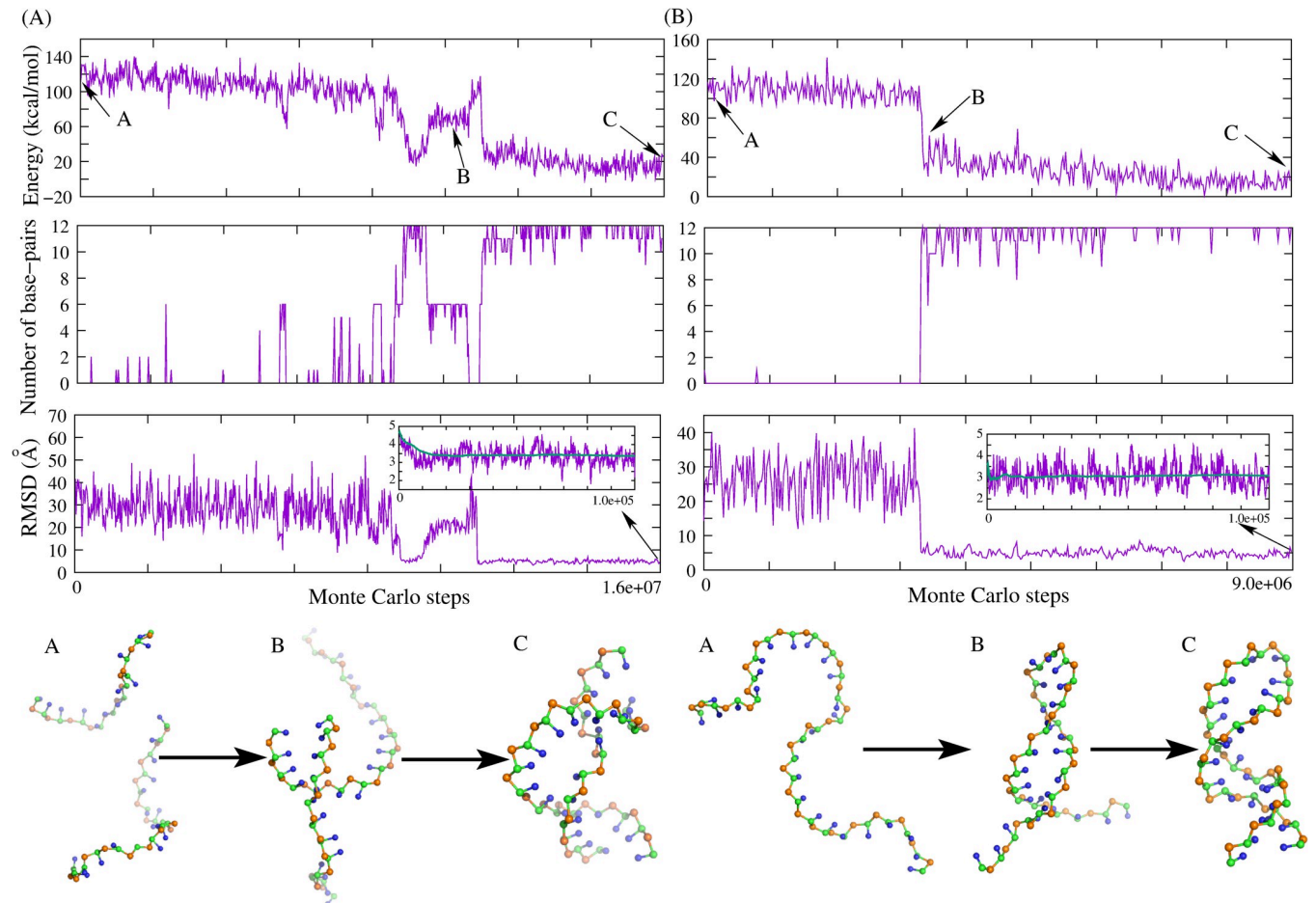
the derivation of which can be found in S1 Text.

## Results

Based on the parameterized implicit-solvent/salt energy function and the MC simulated annealing algorithm, the present CG model can be used to predict 3D structures for dsDNA as well as ssDNA at different ion conditions and temperatures from the sequence. In this section, we tested the present model on the 3D structure and stability predictions for extensive DNAs with various lengths/sequences. As compared with the experimental structures and thermodynamics data, the present model can make overall reliable predictions.

### DNA 3D structure prediction from sequence

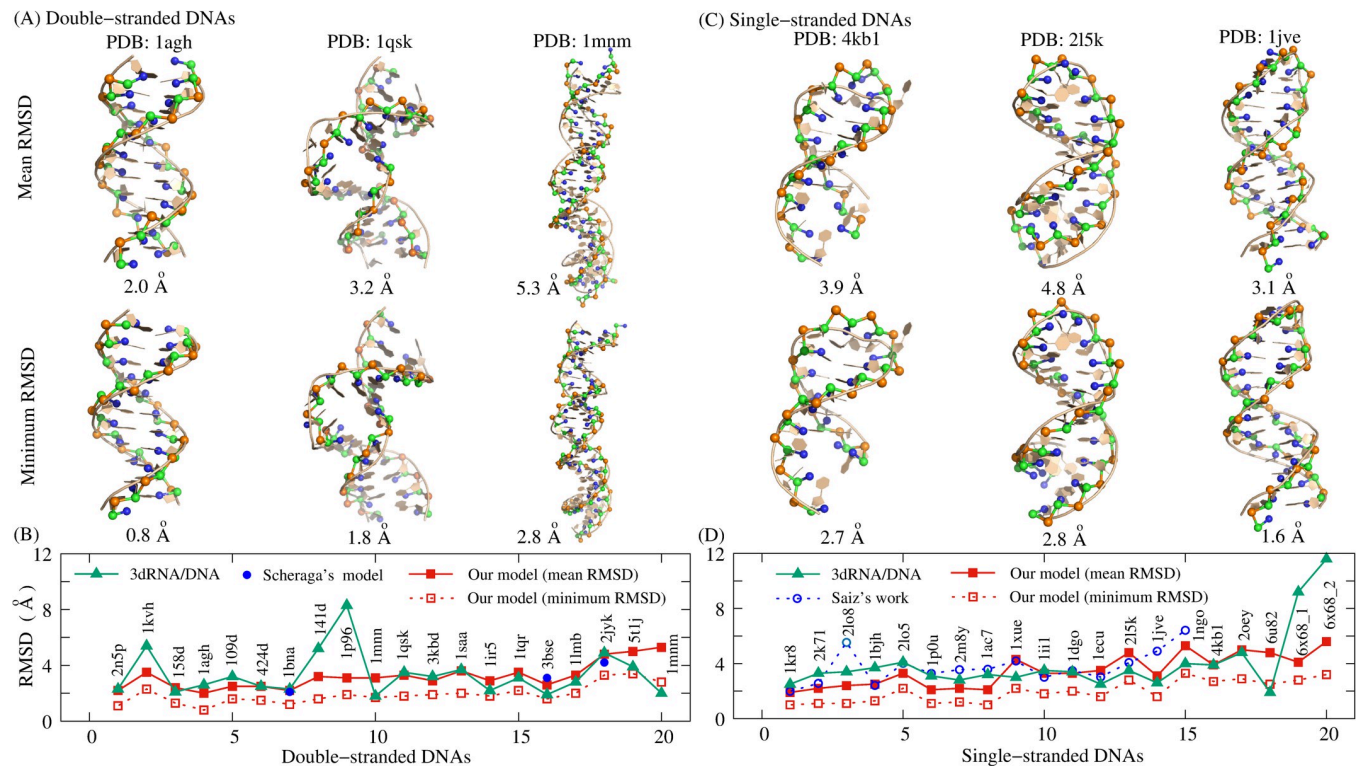
**For dsDNAs.** As described in the section of “Material and methods”, for each dsDNA, two random chains were generated from its sequence (e.g., structure A in Fig 2A), which were further randomly placed in a cubic box, ensuring that there is no overlap. To guarantee no



**Fig 2. 3D structure prediction for the paradigm dsDNA/ssDNA in the present model.** (A,B) The time-evolution of system energy, number of base-pairs, RMSD from native structure, and typical 3D conformations (from top to bottom, respectively) during the Monte Carlo simulated annealing simulation for (A) a dsDNA (PDB: 1qsk) and (B) an ssDNA (PDB: 1jve). The insets show the RMSDs of refined conformations calculated over all CG beads from the corresponding atoms in native structures. The 3D structures are shown with PyMol (<http://www.pymol.org>).

<https://doi.org/10.1371/journal.pcbi.1010501.g002>

significant effect of the box on 3D structures, the strand concentration was set as 1mM (i.e., box side length of 149 Å) for short dsDNA (<10bp) and 0.1mM for longer ones. Due to the lack of the ion conditions for the experimental structures determined by X-ray crystallography, for simplicity, we only predicted the 3D structures for all DNAs at high ion concentrations (e.g., 1M NaCl), regardless of possible ion effects. As shown in Fig 2A, for a dsDNA with a five-adenine bulge loop (PDB: 1qsk; 29nt, 12bp), the energy of the system reduces with the formation of base pairs as the temperature is gradually decreased from 120°C to 25°C, and the initial random chain folds into its native-like double-stranded structures (e.g., structure C in Fig 2A). Following that, another MC simulation (e.g.,  $1 \times 10^5$  steps) is performed at target temperature based on the final structure predicted by the preceding annealing process, and the two sets of bonded potential parameters  $\text{Para}_{\text{nonhelical}}$  and  $\text{Para}_{\text{helical}}$  are employed respectively for the single-strands/loops and base-pairing regions to better capture the geometry of the helical part. As shown in the inset of the bottom panel of Fig 2A, the mean and minimum RMSDs of the dsDNA between predicted structures and its native structure are  $\sim 3.2\text{\AA}$  and  $\sim 1.8\text{\AA}$ , respectively, and the corresponding predicted 3D structures are as shown in Fig 3A.



**Fig 3. The display of typical predicted 3D structures, and comparisons of RMSDs between the present model and other models.** (A,C) The predicted 3D structures (ball-stick) with the mean (top) and minimum (bottom) RMSDs for (A) three sample dsDNAs (PDBs: 1agh, 1qsk, and 1mmn) and (C) three ssDNAs (PDBs: 4kbl, 215k, and 1jve) from their native structures (cartoon). (B,D) The comparison of the predicted structures between the present model and the existing models including the 3dRNA/DNA and the models from Scheraga's or Saiz's group for (B) 20 dsDNAs and (D) 20 ssDNAs. The results of 3dRNA/DNA are predicted through their online server based on the native secondary structures [53]. The other data is taken from refs. 56 and 59. The 3D structures are shown with PyMol (<http://www.pymol.org>).

<https://doi.org/10.1371/journal.pcbi.1010501.g003>

According to the above process, we employed the present model to predict the 3D structures of 20 dsDNAs (18nt-52nt) including helix with bulge loops, and the detailed descriptions (e.g., sequence, length, and structure feature) of these dsDNAs are listed in Table C in *S1 Text*. For the 20 dsDNAs, the overall mean and minimum RMSD values are  $\sim 3.2$  Å and  $\sim 1.9$  Å, respectively; see Fig 3A and Table C in *S1 Text*, which suggest that the present model can make reliable predictions for 3D structures of dsDNA just from the sequence, despite a certain deviation (especially at the two ends) between the predicted and experimental structures for large dsDNA (e.g., PDB: 1mmn and 5t1j). Fig 3A also shows the predicted 3D structures (ball-stick) with the mean and minimum RMSDs and the experimental structures (cartoon) for three typical dsDNAs with different lengths and sequences, intuitively indicating the ability of the model.

**For ssDNAs.** Compared with most of the existing models, the present model cannot only predict the double helix structure of dsDNA, but it can also make a prediction on the 3D structure for more flexible ssDNA. Similarly, a random chain generated from one ssDNA sequence can fold into native-like structures with temperature dropping; see Fig 2B for an example of a DNA hairpin (PDB: 1jve; 27nt, 12bp), which could primarily benefit from the use of the soft parameters ( $\text{Para}_{\text{nonhelical}}$ ) of bonded potentials and sequence-dependent base-stacking interactions in the present model. As shown in Fig 3D, for 20 ssDNAs (7nt-74nt) used in this work including hairpins with bulge/internal loops (Table D in *S1 Text*), the overall mean (minimum) RMSD between the predicted and experimental structures is  $\sim 3.5$  Å ( $\sim 2.0$  Å), which

strongly suggested that the present model can successfully predict 3D structures for simple ssDNA.

Since the structures of the largest hairpin (i.e., 6x68\_2) from the *piggyBac* DNA transposon (PDB: 6x68, a synaptic protein-DNA complex) has a significant bending possibly influenced by protein [84], our predictions without regard to protein has a certain deviation (mean RMSDs of 5.6 Å) from the experimental structure; see Fig 3D. It is worth noting that beyond DNA hairpins, we also tried to predict the 3D structure for a DNA three-way junction using the present model. As shown in Fig E in S1 Text, the structures (two hairpins at two ends) predicted from the sequence are pretty inconsistent with experimental ones. To find why, we further performed a MC simulation using the present model for the ssDNA starting from its PDB structure, and found that there is no significant difference in energies between predicted and simulated conformations (Fig E in S1 Text), which suggests that some tertiary interactions including the non-canonical base-pairing and base-backbone hydrogen bonding [85] and a more efficient algorithm (e.g., replica-exchanged MC) should be further taken into account in the model.

**Comparisons with other models.** To further examine the ability of the model on predicting 3D structures of DNAs (ssDNA and dsDNA), we also made comparisons with available results from the existing models. First, we employed the 3dRNA/DNA web server (<http://biophy.hust.edu.cn/new/3dRNA/create>), which is an automatic, fast, and high-accuracy RNA and DNA tertiary structure prediction method [53–55], to predict 3D structures for all DNAs used in this work using the default options (e.g., Procedure: best; Loop Building Method: Bi-residue; # of Predictions: 5) and experimental secondary structures, and calculated the mean RMSD of returned conformations for each DNA over the atoms of P, C4' and N1/N9 from the corresponding atoms in the experimental structures. As shown in Fig 3, for 20 dsDNAs, the overall mean RMSD (~3.2 Å) from the present model is not worse than that (~3.3 Å) from the 3dRNA/DNA, and for 20 ssDNAs, our prediction (overall mean RMSD: ~3.5 Å) is slightly smaller than predicted result (~4.0 Å) from the 3dRNA/DNA.

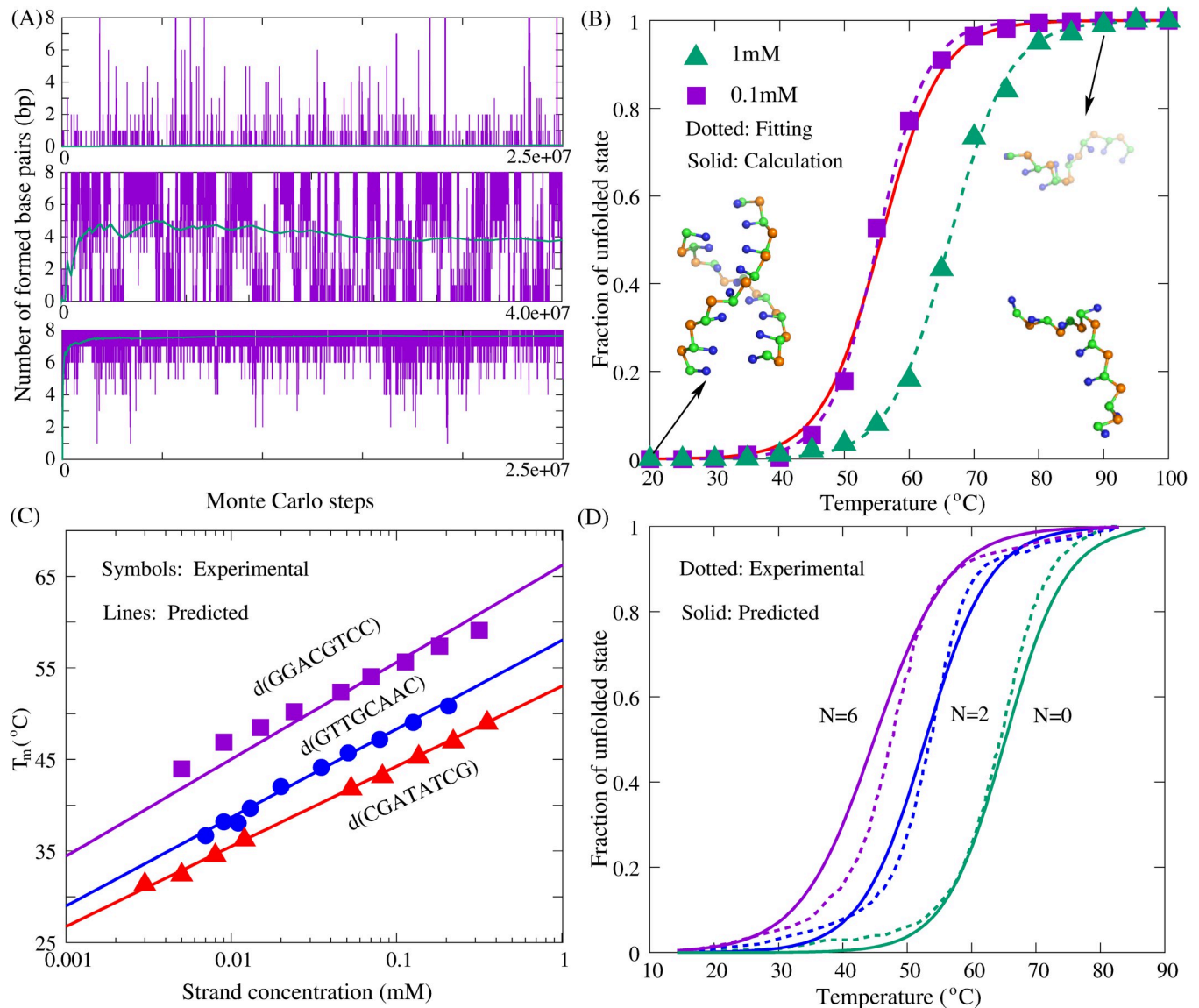
Furthermore, we also made comparisons with the predictions from Refs. 56 and 59. Scheraga et al. also proposed a physics-based rigid-body CG model (3-bead) of DNA, and used it to successfully fold 3 dsDNAs (PDBs: 1bna, 3bse, and 2jyk) from complementary strands with only weak constraints between them [59]. The all-bead RMSDs of the three lowest-energy predicted structures with respect to experimental references are 2.1 Å, 3.1 Å, and 4.2 Å, respectively, which are close to the mean RMSDs (2.2 Å, 2.6 Å, and 4.8 Å, respectively) predicted from the present model (Fig 3A). Jeddi and Saiz presented a pipeline that integrates sequentially building ssDNA secondary structure from Mfold, constructing equivalent 3D ssRNA models by Assemble 2, transforming the 3D ssRNA models into ssDNA 3D structures, and refining the resulting ssDNA 3D structures through MD simulations [56]. As shown in Fig 3D, for 15 ssDNA hairpins, the average RMSD (over the sugar-phosphate backbone) for the best structures predicted by the pipeline is ~3.7 Å, a visibly larger value than the overall mean/minimum RMSD (~3.2 Å/~2.2 Å) from our predictions. Therefore, the comparisons with the other models fully show that the present model can successfully fold simple dsDNA/ssDNA from the sequence without the help of any secondary structure information.

## Stability of various DNAs

Beyond 3D structure predictions, the present model can also be used to predict the thermal stability for dsDNA and ssDNA in ion solutions. In order to verify the effect of the model, we further used it to predict the melt temperatures for extensive DNAs.

**For dsDNA with various lengths/sequences.** The melting temperature ( $T_m$ ) of each dsDNA can be calculated by the present model based on 3D structures predicted at different

temperatures; see Fig 4A and the section “Material and methods”. For example, for the sequence (GGACGTCC)<sub>2</sub> at 1M [Na<sup>+</sup>], the melting curve of the dsDNA with a high strand concentration of 1 mM was predicted according to the fractions of unfolded state at different temperatures (Eqs 5–7), and the melting curve, as well as the  $T_m$  of the dsDNA at low experimental strand concentration (0.1 mM), can be obtained through Eq 8; see Fig 4A and 4B. The predicted  $T_m$  of the sample sequence at  $c_s = 0.1\text{mM}$  is  $\sim 56.0^\circ\text{C}$ , which is only  $0.9^\circ\text{C}$  higher



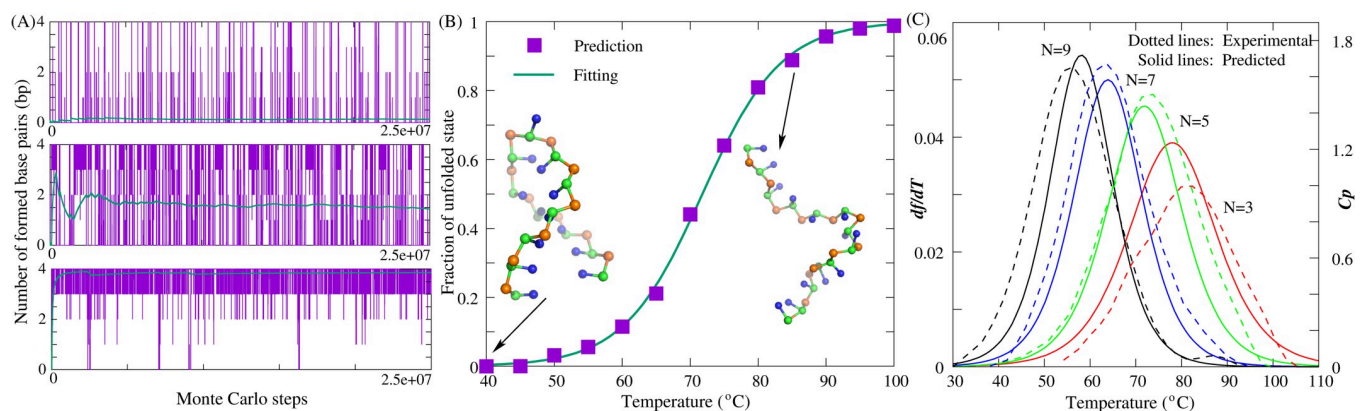
**Fig 4. The stability predictions for dsDNAs in the present model.** (A) The time-evolution of the number of base-pairs for a dsDNA (sequence: (GGACGTCC)<sub>2</sub>; strand concentration: 1mM) at different temperatures (90°C, 65°C, 40°C from top to bottom, respectively) in 1M NaCl solution. (B) The fractions of unfolded state  $f_u$  as functions of temperature for the dsDNA in (A). Green triangle: predictions at high strand concentration (1mM). Purple square: predictions at experimental strand concentration (0.1mM). Two dotted lines are the fitted melting curves to the corresponding predicted data. The solid line is calculated through Eq 7. Ball-stick: the typical 3D structures predicted at low and high temperatures shown with PyMol (<http://www.pymol.org>). (C) The melting temperatures ( $T_m$ 's) as functions of strand concentration for three dsDNAs: purple, (GGACGTCC)<sub>2</sub>, blue, (GTTGCAAC)<sub>2</sub>, and red, (CGATATCG)<sub>2</sub> at 1M [Na<sup>+</sup>]. Symbols: experimental results [71]. Lines: predictions from the present model. (D) The predicted (solid lines) and experimental (dotted lines) melting curves [89] for the dsDNA harboring symmetric internal loops with sequences of CTCGTC(T)<sub>N</sub>CAGTGC/GCACTG(T)<sub>N</sub>GACGAG in 1M NaCl solution. Green: N = 0, i.e., the double helix without internal loop. Blue: N = 2. Purple: N = 6.

<https://doi.org/10.1371/journal.pcbi.1010501.g004>

than the corresponding experimental value (~55.1 °C) [71]. We further performed simulations for the dsDNA at  $c_s = 0.1\text{mM}$  to directly predict its  $T_m$  at experimental strand concentration, and found that there is no significant difference between two melting temperatures, while the melting curve inferred from  $c_s^h$  is slightly broader than that predicted at  $c_s$  (Fig 4B). In addition, as shown in Fig 4C, the predicted  $T_m$ 's for three different dsDNAs at different strand concentrations are also in good accordance with the experiments [71], proving that it is feasible to infer the  $T_m$  at low  $c_s$  from the high ones ( $c_s^h$ ) [82,83].

To examine the sequence effect, 27 dsDNAs (8-36nt) with different sequences have been studied with the present model. The sequences, strand concentrations, and the predicted/experimental melting temperatures are listed in Table 1 [71,86–89]. Here, all dsDNAs are assumed at  $1\text{M} [\text{Na}^+]$  to make comparisons with corresponding experimental data. As shown in Table 1, the  $T_m$  values of extensive dsDNAs from the present model are very close to the experimental measurements with a mean deviation of  $1.5^\circ\text{C}$  and maximal deviations  $< 3.0^\circ\text{C}$ , which indicates that the present model with the sequence-dependent base-stacking/pairing potential can make successful predictions on the stability for dsDNA of extensive sequences/lengths. Furthermore, due to the involvement of coaxial stacking potential, the present model can also provide reliable stability for dsDNA with bulge/internal loops. For example, for 4 dsDNAs with bulge loops and 5 dsDNAs with internal loops, the mean deviation of predicted  $T_m$ 's from the experiments is only  $1.8^\circ\text{C}$ ; see Table 1, and the predicted melting curves for the dsDNAs with internal loops of different lengths are also in line with the experiments [89] (Fig 4D). Fig 4D also shows that the predicted curves of dsDNA with large internal loops (e.g.,  $N = 6$ ) are slightly broader than the experiments, and the possible reason could be that the way of melting temperature calculation used in the present model ignores the difference between melting curves at low and high strand concentrations; see Fig 4D.

**For ssDNA with various lengths/sequences.** Beyond the dsDNA, the stability of ssDNA can also be captured by the present model [36,90–96]. As shown in Fig 5, for DNA hairpins (GCGC(T)<sub>N</sub>GCGC) with different loop lengths ( $N = 3–9$ ), the predicted thermal unfolding curves at  $0.1\text{M} [\text{Na}^+]$  agree reasonably with the experiments, despite that the predicted  $T_m$  (~78 °C) for the hairpin with a small loop (e.g.,  $N = 3$ ) is rather lower than the experimental value (~80.7 °C), while it is a little higher (~58.1 °C vs ~56.2 °C) for large hairpin loops (e.g.,



**Fig 5. The stability predictions for ssDNA hairpins in the present model.** (A) The time-evolution of the number of base-pairs for a simple hairpin (GCGC(T)<sub>5</sub>GCGC) at different temperatures (90 °C, 70 °C, 50 °C from top to bottom, respectively) in 0.1M NaCl solution. (B) The fraction of unfolded state  $f$  as a function of temperature for the hairpin in (A). Symbols: predictions from the present model. Line: fitted melting curve to the predicted data through Eqs 4–6. Ball-stick: the typical 3D structures predicted at low and high temperatures shown with PyMol (<http://www.pymol.org>). (C) The comparisons between predictions (solid lines) and experiments (dotted lines) for four DNA hairpins (GCGC(T)<sub>N</sub>GCGC) with different loop lengths at  $0.1\text{M} [\text{Na}^+]$ . Red:  $N = 3$ . Green:  $N = 5$ . Blue:  $N = 7$ . Black:  $N = 9$ .  $df/dT$ : the first derivative of predicted  $f$  with the temperature.  $C_p$ : the heat capacity from experiment [90].

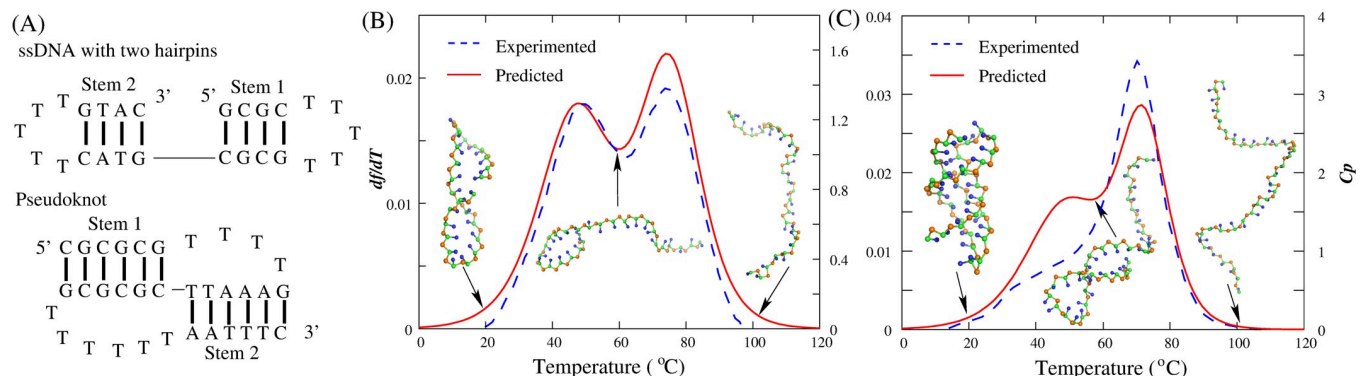
<https://doi.org/10.1371/journal.pcbi.1010501.g005>

$N = 9$ ) [92]. Moreover, 24 ssDNAs including pseudoknot are used to verify the ability of the present model for sequence effect on stability; see Table 2. In order to compare with experiments [36,90–96], all these predictions are at corresponding experimental ion conditions. As shown in Table 2, for 24 ssDNAs with different sequences and lengths (11–34nt), the mean/maximal deviation of  $T_m$  between predicted and experimental is  $\sim 2.1^\circ\text{C}/\sim 3.8^\circ\text{C}$ , which suggests that the effect of sequence on ssDNA stability can also be well described by the present model. It is worth noting that due to the lack of stacking interactions between unpaired bases, the present model cannot distinguish the stability of DNAs with same stems but different loop sequences (e.g.,  $\text{GCGC}(\text{T})_5\text{GCGC}$  vs  $\text{GCGC}(\text{A})_5\text{GCGC}$ ), and yet the stability of them generally differ somewhat [9,91,95].

Specifically, we made additional predictions for the stability of two more complex ssDNAs: a pseudoknot and a chain with two hairpins at two ends; see Fig 6A. As shown in Table 2 and Fig 6, for the ssDNA with two hairpins, two melting temperatures ( $T_{m1}$  and  $T_{m2}$ ) of the corresponding transitions are successfully predicted by the present model, with the deviations of  $\sim 2.1^\circ\text{C}$  and  $\sim 1.1^\circ\text{C}$  from experimental data, respectively. Since the hairpin at 3' end contains fewer G-C pairs than the other (Fig 6A), it melts at a significantly low temperature in comparison to the 5' end hairpin [92]. For the DNA pseudoknot at  $0.1\text{M} [\text{Na}^+]$ , the predicted  $T_{m1}$  and  $T_{m2}$  are  $\sim 48.8^\circ\text{C}$  and  $\sim 72.0^\circ\text{C}$ , respectively, which also agree well with the experimental data ( $\sim 52.6^\circ\text{C}$  and  $\sim 70.7^\circ\text{C}$ ) [90]; see Table 2, and the comparison between predicted and experimental thermal unfolding curves can be found in Fig 6C. In the predicted curve, the first transition that is from folded pseudoknot state to intermediate hairpin state is more significant than that from experiment. One possible reason is that noncanonical interactions such as triple base interactions between loops and stems and self-stacking in loop nucleotides, which are common in RNA/DNA pseudoknots [69,90], are neglected by the present model, leading to a relatively simple unfolding energy surface. Even so, the comparison with the experiment still suggests that the present model can be reliable in predicting thermal stability for DNA pseudoknots in monovalent ion solutions, and it is noted that the present model can also provide 3D structures for the pseudoknot at different temperatures from the sequences.

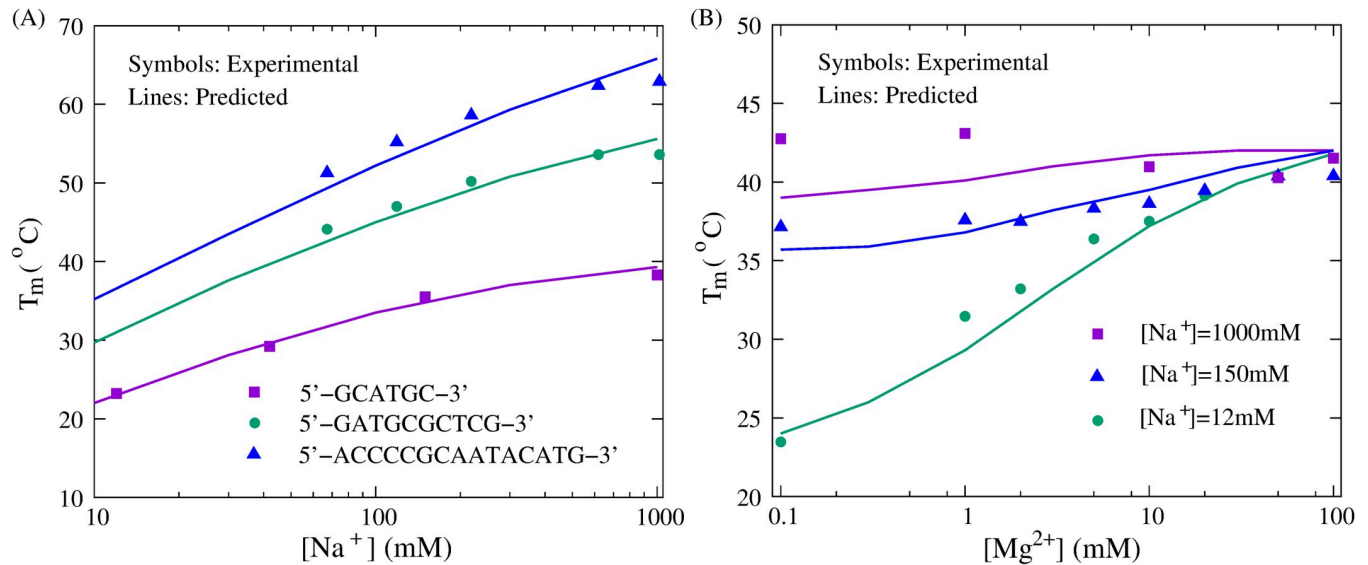
### Monovalent/Divalent ion effects on stability of dsDNA/ssDNA

Due to the high density of negative charges on the backbone, DNA stability is sensitive to the ionic condition of the solution, while the effect of ions, especially divalent ions (e.g.,  $\text{Mg}^{2+}$ ), is



**Fig 6. The stability prediction for an ssDNA with two hairpins and a DNA pseudoknot in the present model.** (A) The schematics of secondary structure for the ssDNA with two hairpins (top) and the DNA pseudoknot (bottom). (B,C) The comparisons between predictions (solid lines) and experiments (dotted lines) for the two ssDNAs in (a).  $df/dT$ : the first derivative of predicted  $f$  with the temperature.  $C_p$ : the heat capacity from experiments [90,92]. Ball-stick: the typical 3D structures predicted at different temperatures shown with PyMol (<http://www.pymol.org>).

<https://doi.org/10.1371/journal.pcbi.1010501.g006>



**Fig 7. The comparisons of stability between predictions (lines) and experiments (symbols) for dsDNAs in monovalent/divalent ion solutions.** (A) The melting temperatures ( $T_m$ 's) as functions of  $[Na^+]$  for three dsDNAs with sequences of (GCATGC)<sub>2</sub>, (GATGCGCTCG)<sub>2</sub>, (ACCCGCAATACATG)<sub>2</sub> from bottom to top, respectively. Symbols: experimental data [87,97]. Lines: predictions from the present model. (B) The melting temperatures ( $T_m$ 's) as functions of  $[Mg^{2+}]$  for the dsDNA with sequence of (GCATGC)<sub>2</sub> at different  $[Na^+]$ 's: 0.012M, 0.15M, and 1M from bottom to top, respectively. Symbols: experimental data [87]. Lines: predictions from the present model.

<https://doi.org/10.1371/journal.pcbi.1010501.g007>

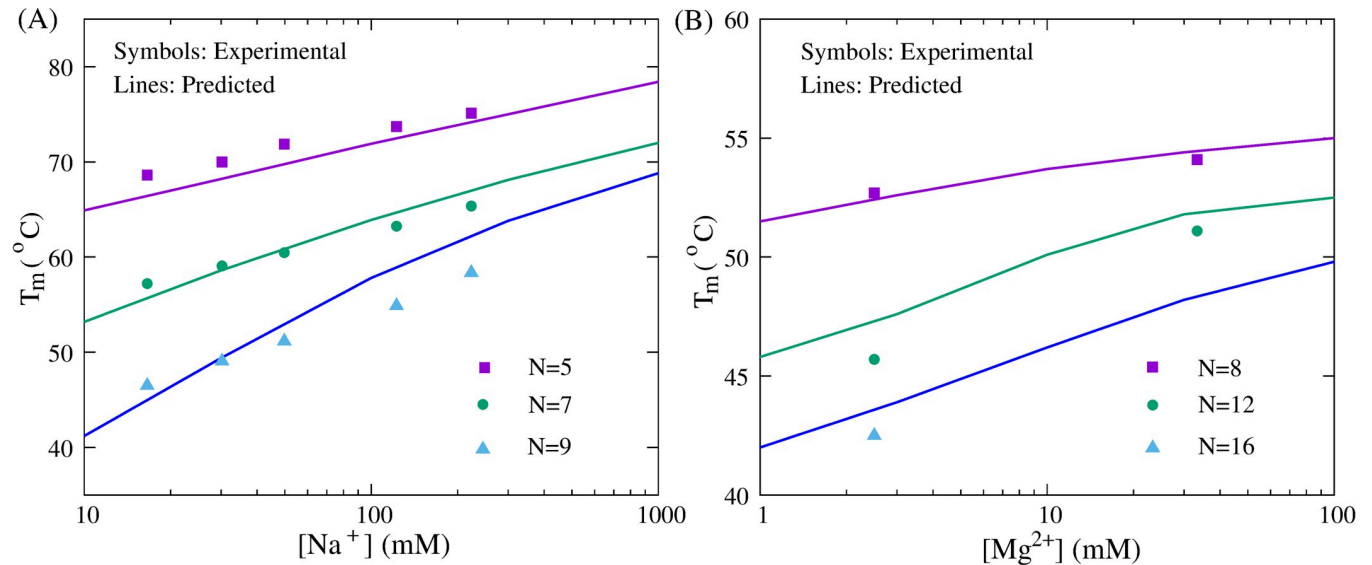
generally ignored in the existing DNA CG models [43–50]. Here, we employed the present model to examine the monovalent/divalent ion effects on the thermal stability of dsDNA and ssDNA.

**Monovalent ion effect.** For each of the three dsDNAs with different lengths (6bp, 10bp, and 15bp), we performed simulations over a broad range of monovalent ion concentrations ( $[Na^+]$ : 0.01M–1.0M), and calculated the melting temperatures at different  $[Na^+]$ 's. As shown in Fig 7A, the increase of  $[Na^+]$  enhances the dsDNA folding stability due to the stronger ion neutralization [62,63], and the predicted melting temperatures for the three dsDNAs are well in accordance with the experiment results [87,97], with the mean deviation of  $\sim 1.4^\circ C$ . Fig 7A also shows that the  $[Na^+]$ -dependence of  $T_m$  is stronger for longer dsDNA, which could be caused by the larger buildup of negative charges during base pair formation of longer dsDNA [63,76].

Although Table 2 has indicated that the present model can make reliable predictions for ssDNA stability at various  $[Na^+]$ 's, we further used a simple DNA hairpin (GCGC(T)<sub>N</sub>GCGC) with different loop lengths ( $N = 5, 7,$  and  $9$ ) to test monovalent ion effect on stability in the present model. As shown in Fig 8A, for the hairpin with small loops (e.g.,  $N = 5$  and  $7$ ), the difference of predicted  $T_m$  from the experiments over a wide range of  $[Na^+]$ 's is very small (e.g., mean/maximal deviation of  $\sim 1.5^\circ C/\sim 1.0^\circ C$ ), and for the loop length of  $9$ , our predictions are slightly larger than the experimental data only at high  $[Na^+]$ 's; e.g.,  $\sim 4.0^\circ C$  higher at  $\sim 0.1M$   $[Na^+]$  [90]. The results on the stability of ssDNA and dsDNA in monovalent ion solutions reveal that it is a very effective way of involving the electrostatic interaction for DNA in the present model through the combination of the Debye-Huckel approximation and the concept of counterion condensation, which has also been validated by the TIS model [50,51].

**Divalent ion effect.** Remarkably, one important feature of the present model is that combining the counterion condensation theory and the results from the TBI model; see Eq 3, it can also be used to simulate DNA folding in mixed monovalent/divalent ion solutions. For one





**Fig 8. The comparisons of stability between predictions (lines) and experiments (symbols) for ssDNAs in monovalent/divalent ion solutions.** (A) The melting temperatures ( $T_m$ 's) as functions of  $[\text{Na}^+]$  for the DNA hairpins (GCGC(T)<sub>N</sub>GCGC) with different loop lengths: 5, 7, and 9 from top to bottom, respectively. Symbols: experimental data [90]. Lines: predictions from the present model. (B) The melting temperatures ( $T_m$ 's) as functions of  $[\text{Mg}^{2+}]$  for the DNA hairpins (CGGATAA(T)<sub>N</sub>TTATCCG) with different loop lengths: 8, 12, and 16 from top to bottom, respectively. Symbols: experimental data at 2.5mM or 33mM  $[\text{Mg}^{2+}]$  with 10mM Tris-HCl buffer [96]. Lines: predictions from the present model at extensive  $[\text{Mg}^{2+}]$ 's without  $\text{Na}^+$ .

<https://doi.org/10.1371/journal.pcbi.1010501.g008>

dsDNA with the sequence of (GCATGC)<sub>2</sub> and one ssDNA hairpin with various lengths of the loop (CGGATAA(T)<sub>N</sub>TTATCCG), we made massive predictions in mixed  $\text{Na}^+/\text{Mg}^{2+}$  solutions and compared the melting temperatures with the corresponding experimental results [87,96]; see Figs 7B and 8B. The comparisons of  $T_m$ 's over a wide range of  $[\text{Mg}^{2+}]$  are in line with the experiments, whether for the dsDNA at different  $[\text{Na}^+]$ 's (i.e., 0.012M, 0.15M, and 1M) or for the ssDNA with different lengths of the loop, which suggests that the present model can nearly make quantitative predictions for the stability of DNAs in mixed ion solutions from their sequences, even though the ion effect is involved implicitly.

Furthermore, the competition between  $\text{Na}^+$  and  $\text{Mg}^{2+}$  on DNA stability can also be captured by the present model. For example, for dsDNA at 0.012M  $[\text{Na}^+]$  (Fig 7B), when  $[\text{Mg}^{2+}]$  is very low (e.g., <0.3mM),  $\text{Na}^+$  dominates the stability of the dsDNA, while the increase of  $[\text{Mg}^{2+}]$  enhances the stability significantly. This is because the bindings of  $\text{Na}^+$  and  $\text{Mg}^{2+}$  are generally anti-cooperative and  $\text{Mg}^{2+}$ -binding is more efficient in stabilizing DNA structures [63,75]. Naturally, as  $[\text{Na}^+]$  increases, the negative charge on DNA is strongly neutralized, and consequently, the effect of  $\text{Mg}^{2+}$  appears weak. In particular, as shown in Fig 7B, there is a significant deviation between predicted and experimental  $T_m$ 's for the dsDNA at 1M  $[\text{Na}^+]$  and various  $[\text{Mg}^{2+}]$ 's. The possible reason could be that when the ion concentration is high enough (e.g., >1M  $[\text{Na}^+]$ ), the effect of electrostatic interaction on DNA stability is quite negligible, and the competition between  $\text{Na}^+$  and  $\text{Mg}^{2+}$  could be dominated by the entropy changes of ions, which is difficult to be precisely described by the implicit ion model used in the present model [62,67,76].

## Discussion

In this work, we have proposed a novel three-bead CG model to predict 3D structure and stability for both ssDNA and dsDNA in ion solutions only from the sequence. As compared with the extensive experiments, we have demonstrated that, (1) The present model can successfully

predict the native-like 3D structures for ssDNAs and dsDNAs with an overall mean (minimum) RMSD of  $\sim 3.4\text{\AA}$  ( $\sim 1.9\text{\AA}$ ) from corresponding experimental structures, and the overall prediction accuracy of the present model is slightly higher than the existing models; (2) The present model can make reliable predictions on stability for dsDNAs with/without bulge loops and ssDNAs including pseudoknots, and for 51 DNAs with various lengths and sequences, the predicted melting temperatures are in good accordance with extensive experiment data (i.e., mean deviation of  $\sim 2.0^\circ\text{C}$ ); (3) The present model with implicit electrostatic potential can also reproduce the stability for ssDNAs/dsDNAs at extensive monovalent or mixed monovalent/divalent ion conditions, with the predicted melting temperatures consistent with the available experiments.

Nonetheless, the present model has several limitations that should be overcome in future model development. For example, the present model failed to predict native-like structures for more complex DNAs such as that with triplexes, quadruplexes or n-way junction and cannot distinguish the stability for DNAs with different loop sequences, which suggest that possible noncanonical interactions (e.g., noncanonical base-pairing, base triple interactions between loops and stems, self-stacking in loop nucleotides and special hydrogen bonds involving phosphates and sugars) should be further taken into account [2,50,85]. Furthermore, a more efficient sampling algorithm such as replica-exchanged MC or MC with umbrella sampling, as well as suitable structure constraints should be introduced to the model for assembly of large DNAs (e.g., nano-architectures) [46–49,50], and accordingly, an accurate score function like statistical potential used for RNA and protein could be required to evaluate predicted DNA candidate structures [98–103]. In addition, the 3D structure predicted by the present model is at the CG level, and it is still necessary to reconstruct all-atomistic structures based on the CG structures for further practical applications. After these further developments, a user-friendly web server would be further freely available, allowing users to predict 3D structure and stability for DNAs in ion solutions from sequence or given constraints.

## Supporting information

**S1 Text.** The force field of the present model, the melting temperature calculation for dsDNAs at low strand concentrations, and the additional figures (Figs A-E) and tables (Tables A-D). (PDF)

## Acknowledgments

We are grateful to Profs. Zhi-Jie Tan (Wuhan University), Wenbing Zhang (Wuhan University), and Yaoqi Zhou (Shenzhen Bay Laboratory) for valuable discussions, and we would like to acknowledge computing resources from the Shenzhen Bay Laboratory Supercomputing Center.

## Author Contributions

**Conceptualization:** Zi-Chun Mu, Ya-Zhou Shi.

**Data curation:** Zi-Chun Mu, Ya-Lan Tan.

**Formal analysis:** Zi-Chun Mu, Ya-Lan Tan, Ya-Zhou Shi.

**Funding acquisition:** Ben-Gong Zhang, Ya-Zhou Shi.

**Investigation:** Jie Liu, Ya-Zhou Shi.

**Methodology:** Zi-Chun Mu, Ya-Lan Tan, Ben-Gong Zhang, Ya-Zhou Shi.

**Supervision:** Ya-Zhou Shi.

**Validation:** Zi-Chun Mu, Ya-Lan Tan, Ya-Zhou Shi.

**Visualization:** Zi-Chun Mu.

**Writing – original draft:** Zi-Chun Mu, Ya-Zhou Shi.

**Writing – review & editing:** Zi-Chun Mu, Jie Liu, Ya-Zhou Shi.

## References

1. Guiblet WM, Cremona MA, Harris RS, Chen D, Eckert KA, Chiaromonte F, et al. Non-B DNA: a major contributor to small- and large-scale variation in nucleotide substitution frequencies across the genome. *Nucleic Acids Res.* 2021; 49: 1497–1516. <https://doi.org/10.1093/nar/gkaa1269> PMID: 33450015
2. Ferry G. The structure of DNA. *Nature.* 2019; 575: 35–36. <https://doi.org/10.1038/d41586-019-02554-z> PMID: 31686042
3. Robinson J, Raguseo F, Nuccio SP, Liano D, Antonio DM. DNA G-quadruplex structures: more than simple roadblocks to transcription? *Nucleic Acids Res.* 2021; 49: 8419–8431. <https://doi.org/10.1093/nar/gkab609> PMID: 34255847
4. Varshney D, Spiegel J, Zyner K, Tannahill D, Balasubramanian S. The regulation and functions of DNA and RNA G-quadruplexes. *Nat Rev Mol Cell Biol.* 2020; 21: 459–474. <https://doi.org/10.1038/s41580-020-0236-x> PMID: 32313204
5. Seeman N, Sleiman H. DNA nanotechnology. *Nat Rev Mater.* 2018; 3: 17068.
6. Hu Q, Li H, Wang L, Gu H, Fan C. DNA nanotechnology-enabled drug delivery systems. *Chem Rev.* 2019; 119: 6459–6506.
7. Dey S, Fan C, Gothelf KV, Li J, Lin C, Liu L, et al. DNA origami. *Nat Rev Methods Primers.* 2021, 1: 13.
8. Ma W, Zhan Y, Zhang Y, Mao C, Xie X, Lin Y. The biological applications of DNA nanomaterials: current challenges and future directions. *Signal Transduct Target Ther.* 2021; 6: 351. <https://doi.org/10.1038/s41392-021-00727-9> PMID: 34620843
9. SantaLucia J Jr., Hicks D. The thermodynamics of DNA structural motifs. *Annu Rev Biophys Biomol Struct.* 2004; 33: 415–440. <https://doi.org/10.1146/annurev.biophys.32.110601.141800> PMID: 15139820
10. Bryant Z, Oberstrass FC, Basu A. Recent developments in single-molecule DNA mechanics. *Curr Opin Struct Biol.* 2012; 22: 304–312. <https://doi.org/10.1016/j.sbi.2012.04.007> PMID: 22658779
11. Kriegel F, Ermann N, Lipfert J. Probing the mechanical properties, conformational changes, and interactions of nucleic acids with magnetic tweezers. *J Struct Biol.* 2017; 197: 26–36. <https://doi.org/10.1016/j.jsb.2016.06.022> PMID: 27368129
12. Cruz-Leon S, Vanderlinden W, Muller P, Forster T, Staudt G, Lin YY, et al. Twisting DNA by salt. *Nucleic Acids Res.* 2022; 50: 5726–5738. <https://doi.org/10.1093/nar/gkac445> PMID: 35640616
13. Fu H, Zhang C, Qiang XW, Yang YJ, Dai L, Tan ZJ, et al. Opposite effects of high-valent cations on the elasticities of DNA and RNA duplexes revealed by magnetic tweezers. *Phys Rev Lett.* 2020; 124: 058101. <https://doi.org/10.1103/PhysRevLett.124.058101> PMID: 32083903
14. Zhang Y, Zhou H, Ou-Yang ZC. Stretching single-stranded DNA: interplay of electrostatic, base-pairing, and base-pair stacking interactions. *Biophys J.* 2001; 81: 1133–1143. [https://doi.org/10.1016/S0006-3495\(01\)75770-0](https://doi.org/10.1016/S0006-3495(01)75770-0) PMID: 11463654
15. Di W, Gao X, Huang W, Sun Y, Lei H, Liu Y, et al. Direct measurement of length scale dependence of the hydrophobic free energy of a single collapsed polymer nanosphere. *Phys Rev Lett.* 2019; 122: 047801. <https://doi.org/10.1103/PhysRevLett.122.047801> PMID: 30768307
16. Minhas V, Sun T, Mirzoev A, Korolev N, Lyubartsev AP, Nordenskiöld L. Modeling DNA flexibility: comparison of force fields from atomistic to multiscale levels. *J Phys Chem B.* 2020; 124: 38–49. <https://doi.org/10.1021/acs.jpcc.9b09106> PMID: 31805230
17. Liebl K, Zacharias M. Accurate modeling of DNA conformational flexibility by a multivariate Ising model. *Proc Natl Acad Sci U S A.* 2021; 118:e2021263118. <https://doi.org/10.1073/pnas.2021263118> PMID: 33876759
18. Jones MS, Ashwood B, Tokmakoff A, Ferguson AL. Determining sequence-dependent DNA oligonucleotide hybridization and dehybridization mechanisms using coarse-grained molecular simulation, markov state models, and infrared spectroscopy. *J Am Chem Soc.* 2021; 143: 17395–17411. <https://doi.org/10.1021/jacs.1c05219> PMID: 34644072

19. He J, Wang J, Tao H, Xiao Y, Huang SY. HNADOCK: a nucleic acid docking server for modeling RNA/DNA-RNA/DNA 3D complex structures. *Nucleic Acids Res.* 2019; 47: W35–W42. <https://doi.org/10.1093/nar/gkz412> PMID: 31114906
20. Bao L, Zhang X, Shi YZ, Wu YY, Tan ZJ. Understanding the relative flexibility of RNA and DNA duplexes: stretching and twist-stretch coupling. *Biophys J.* 2017; 112: 1094–1104. <https://doi.org/10.1016/j.bpj.2017.02.022> PMID: 28355538
21. Wu YY, Bao L, Zhang X, Tan ZJ. Flexibility of short DNA helices with finite-length effect: From base pairs to tens of base pairs. *J Chem Phys.* 2015; 142: 125103. <https://doi.org/10.1063/1.4915539> PMID: 25833610
22. Wang Y, Gong S, Wang Z, Zhang W. The thermodynamics and kinetics of a nucleotide base pair. *J Chem Phys.* 2016; 144: 115101. <https://doi.org/10.1063/1.4944067> PMID: 27004898
23. Liu T, Yu T, Zhang S, Wang Y, Zhang W. Thermodynamic and kinetic properties of a single base pair in A-DNA and B-DNA. *Phys Rev E.* 2021; 103: 042409. <https://doi.org/10.1103/PhysRevE.103.042409> PMID: 34005973
24. Kriegel F, Matek C, Drsata T, Kulenkampff K, Tschirpke S, Zacharias M, et al. The temperature dependence of the helical twist of DNA. *Nucleic Acids Res.* 2018; 46: 7998–8009. <https://doi.org/10.1093/nar/gky599> PMID: 30053087
25. Zerze GH, Stillinger FH, Debenedetti PG. Thermodynamics of DNA hybridization from atomistic simulations. *J Phys Chem B.* 2021; 125: 771–779. <https://doi.org/10.1021/acs.jpcc.0c09237> PMID: 33434025
26. Upadhyaya A, Kumar S. Effect of loop sequence on unzipping of short DNA hairpins. *Phys Rev E.* 2021; 103: 062411. <https://doi.org/10.1103/PhysRevE.103.062411> PMID: 34271739
27. Nomidis SK, Kriegel F, Vanderlinden W, Lipfert J, Carlon E. Twist-bend coupling and the torsional response of double-stranded DNA. *Phys Rev Lett.* 2017; 118: 217801. <https://doi.org/10.1103/PhysRevLett.118.217801> PMID: 28598642
28. Marko JF, Siggia ED. Fluctuations and supercoiling of DNA. *Science.* 1994; 265: 506–508. <https://doi.org/10.1126/science.8036491> PMID: 8036491
29. Nomidis SK, Skoruppa E, Carlon E, Marko JF. Twist-bend coupling and the statistical mechanics of the twistable wormlike-chain model of DNA: Perturbation theory and beyond. *Phys Rev E.* 2019; 99: 032414. <https://doi.org/10.1103/PhysRevE.99.032414> PMID: 30999490
30. Toan NM, Thirumalai D. On the origin of the unusual behavior in the stretching of single-stranded DNA. *J Chem Phys.* 2012; 136: 235103. <https://doi.org/10.1063/1.4729371> PMID: 22779622
31. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 2003; 31: 3406–3415. <https://doi.org/10.1093/nar/gkg595> PMID: 12824337
32. Ingolfsson HI, Lopez CA, Uusitalo JJ, de Jong DH, Gopal SM, Periole X, et al. The power of coarse graining in biomolecular simulations. *Wiley Interdiscip Rev Comput Mol Sci.* 2014; 4: 225–248. <https://doi.org/10.1002/wcms.1169> PMID: 25309628
33. Dans PD, Walther J, Gomez H, Orozco M. Multiscale simulation of DNA. *Curr Opin Struct Biol.* 2016; 37: 29–45. <https://doi.org/10.1016/j.sbi.2015.11.011> PMID: 26708341
34. Sun T, Minhas V, Korolev N, Mirzoev A, Lyubartsev AP, Nordenskiold L. Bottom-up coarse-grained modeling of DNA. *Front Mol Biosci.* 2021; 8: 645527. <https://doi.org/10.3389/fmolb.2021.645527> PMID: 33816559
35. Reshetnikov RV, Stolyarova AV, Zalevsky AO, Panteleev DY, Pavlova GV, Klinov DV, et al. A coarse-grained model for DNA origami. *Nucleic Acids Res.* 2018; 46: 1102–1112. <https://doi.org/10.1093/nar/gkx1262> PMID: 29267876
36. Linak MC, Dorfman KD. Analysis of a DNA simulation model through hairpin melting experiments. *J Chem Phys.* 2010; 133: 125101. <https://doi.org/10.1063/1.3480685> PMID: 20886965
37. Walther J, Dans PD, Balaceanu A, Hospital A, Bayarri G, Orozco M. A multi-modal coarse grained model of DNA flexibility mappable to the atomistic level. *Nucleic Acids Res.* 2020; 48: e29. <https://doi.org/10.1093/nar/gkaa015> PMID: 31956910
38. Maffeo C, Ngo TT, Ha T, Aksimentiev A. A coarse-grained model of unstructured single-stranded DNA derived from atomistic simulation and single-molecule experiment. *J Chem Theory Comput.* 2014; 10: 2891–2896. <https://doi.org/10.1021/ct500193u> PMID: 25136266
39. Dans PD, Zeida A, Machado MR, Pantano S. A coarse grained model for atomic-detailed DNA simulations with explicit electrostatics. *J Chem Theory Comput.* 2010; 6: 1711–1725. <https://doi.org/10.1021/ct900653p> PMID: 26615701
40. Uusitalo JJ, Ingolfsson HI, Akhshi P, Tieleman DP, Marrink SJ. Martini coarse-grained force field: Extension to DNA. *J Chem Theory Comput.* 2015; 11: 3932–3945. <https://doi.org/10.1021/acs.jctc.5b00286> PMID: 26574472

41. Souza PCT, Alessandri R, Barnoud J, Thallmair S, Faustino I, Grunewald F, et al. Martini 3: a general purpose force field for coarse-grained molecular dynamics. *Nat Methods*. 2021; 18: 382–388. <https://doi.org/10.1038/s41592-021-01098-3> PMID: 33782607
42. Assenza S, Perez R. Accurate sequence-dependent coarse-grained model for conformational and elastic properties of double-stranded DNA. *J Chem Theory Comput*. 2022; 18: 3239–3256. <https://doi.org/10.1021/acs.jctc.2c00138> PMID: 35394775
43. Hinckley DM, Freeman GS, Whitmer JK, de Pablo JJ. An experimentally-informed coarse-grained 3-Site-Per-Nucleotide model of DNA: structure, thermodynamics, and dynamics of hybridization. *J Chem Phys*. 2013; 139: 144903. <https://doi.org/10.1063/1.4822042> PMID: 24116642
44. Knotts TAT, Rathore N, Schwartz DC, de Pablo JJ. A coarse grain model for DNA. *J Chem Phys*. 2007; 126: 084901. <https://doi.org/10.1063/1.2431804> PMID: 17343470
45. Freeman GS, Hinckley DM, Lequeieu JP, Whitmer JK, de Pablo JJ. Coarse-grained modeling of DNA curvature. *J Chem Phys*. 2014; 141: 165103. <https://doi.org/10.1063/1.4897649> PMID: 25362344
46. Ouldridge TE, Louis AA, Doye JP. Structural, mechanical, and thermodynamic properties of a coarse-grained DNA model. *J Chem Phys*. 2011; 134: 085101. <https://doi.org/10.1063/1.3552946> PMID: 21361556
47. Hong F, Schreck JS, Sulc P. Understanding DNA interactions in crowded environments with a coarse-grained model. *Nucleic Acids Res*. 2020; 48: 10726–10738. <https://doi.org/10.1093/nar/gkaa854> PMID: 33045749
48. Poppleton E, Romero R, Mallya A, Rovigatti L, Sulc P. OxDNA.org: a public webserver for coarse-grained simulations of DNA and RNA nanostructures. *Nucleic Acids Res*. 2021; 49: W491–W8. <https://doi.org/10.1093/nar/gkab324> PMID: 34009383
49. Ouldridge TE, Louis AA, Doye JP. DNA nanotweezers studied with a coarse-grained model of DNA. *Phys Rev Lett*. 2010; 104: 178101. <https://doi.org/10.1103/PhysRevLett.104.178101> PMID: 20482144
50. Chakraborty D, Hori N, Thirumalai D. Sequence-dependent three interaction site model for single- and double-stranded DNA. *J Chem Theory Comput*. 2018; 14: 3763–3779. <https://doi.org/10.1021/acs.jctc.8b00091> PMID: 29870236
51. Denesyuk NA, Thirumalai D. Coarse-grained model for predicting RNA folding thermodynamics. *J Phys Chem B*. 2013; 117: 4901–4911. <https://doi.org/10.1021/jp401087x> PMID: 23527587
52. Hyeon C, Thirumalai D. Capturing the essence of folding and functions of biomolecules using coarse-grained models. *Nat Commun*. 2011; 2: 487. <https://doi.org/10.1038/ncomms1481> PMID: 21952221
53. Zhang Y, Xiong Y, Xiao Y. 3dDNA: A computational method of building DNA 3D structures. *Molecules*. 2022; 27: 5936. <https://doi.org/10.3390/molecules27185936> PMID: 36144680
54. Zhang Y, Wang J, Xiao Y. 3dRNA: 3D structure prediction from linear to circular RNAs. *J Mol Biol*. 2022; 434: 167452. <https://doi.org/10.1016/j.jmb.2022.167452> PMID: 35662453
55. Zhao Y, Huang Y, Gong Z, Wang Y, Man J, Xiao Y. Automated and fast building of three-dimensional RNA structures. *Sci Rep*. 2012; 2: 734. <https://doi.org/10.1038/srep00734> PMID: 23071898
56. Jeddi I, Saiz L. Three-dimensional modeling of single stranded DNA hairpins for aptamer-based biosensors. *Sci Rep*. 2017; 7: 1178. <https://doi.org/10.1038/s41598-017-01348-5> PMID: 28446765
57. He Y, Maciejczyk M, Oldziej S, Scheraga HA, Liwo A. Mean-field interactions between nucleic-acid-base dipoles can drive the formation of a double helix. *Phys Rev Lett*. 2013; 110: 098101. <https://doi.org/10.1103/PhysRevLett.110.098101> PMID: 23496746
58. He Y, Liwo A, Scheraga HA. Optimization of a Nucleic Acids united-RESidue 2-Point model (NARES-2P) with a maximum-likelihood approach. *J Chem Phys*. 2015; 143: 243111. <https://doi.org/10.1063/1.4932082> PMID: 26723596
59. Maciejczyk M, Spasic A, Liwo A, Scheraga HA. DNA duplex formation with a coarse-grained model. *J Chem Theory Comput*. 2014; 10: 5020–5035. <https://doi.org/10.1021/ct4006689> PMID: 25400520
60. Cragolini T, Derreumaux P, Pasquali S. Coarse-grained simulations of RNA and DNA duplexes. *J Phys Chem B*. 2013; 117: 8047–8060. <https://doi.org/10.1021/jp400786b> PMID: 23730911
61. Pasquali S, Derreumaux P. HiRE-RNA: a high resolution coarse-grained energy model for RNA. *J Phys Chem B*. 2010; 114: 11957–11966. <https://doi.org/10.1021/jp102497y> PMID: 20795690
62. Lipfert J, Doniach S, Das R, Herschlag D. Understanding nucleic acid-ion interactions. *Annu Rev Biochem*. 2014; 83: 813–841. <https://doi.org/10.1146/annurev-biochem-060409-092720> PMID: 24606136
63. Tan ZJ, Chen SJ. Electrostatic free energy landscapes for nucleic acid helix assembly. *Nucleic Acids Res*. 2006; 34: 6629–6639. <https://doi.org/10.1093/nar/gkl810> PMID: 17145719

64. Wu YY, Zhang ZL, Zhang JS, Zhu XL, Tan ZJ. Multivalent ion-mediated nucleic acid helix-helix interactions: RNA versus DNA. *Nucleic Acids Res.* 2015; 43: 6156–6165. <https://doi.org/10.1093/nar/gkv570> PMID: 26019178
65. Zhang C, Tian F, Lu Y, Yuan B, Tan ZJ, Zhang XH, et al. Twist-diameter coupling drives DNA twist changes with salt and temperature. *Sci Adv.* 2022; 8: eabn1384. <https://doi.org/10.1126/sciadv.abn1384> PMID: 35319990
66. Shi YZ, Wang FH, Wu YY, Tan ZJ. A coarse-grained model with implicit salt for RNAs: predicting 3D structure, stability and salt effect. *J Chem Phys.* 2014; 141: 105102. <https://doi.org/10.1063/1.4894752> PMID: 25217954
67. Shi YZ, Jin L, Wang FH, Zhu XL, Tan ZJ. Predicting 3D Structure, Flexibility, and Stability of RNA Hairpins in Monovalent and Divalent Ion Solutions. *Biophys J.* 2015; 109: 2654–2665. <https://doi.org/10.1016/j.bpj.2015.11.006> PMID: 26682822
68. Jin L, Tan YL, Wu Y, Wang X, Shi YZ, Tan ZJ. Structure folding of RNA kissing complexes in salt solutions: predicting 3D structure, stability, and folding pathway. *RNA.* 2019; 25: 1532–1548. <https://doi.org/10.1261/ma.071662.119> PMID: 31391217
69. Shi YZ, Jin L, Feng CJ, Tan YL, Tan ZJ. Predicting 3D structure and stability of RNA pseudoknots in monovalent and divalent ion solutions. *PLoS Comput Biol.* 2018; 14: e1006222. <https://doi.org/10.1371/journal.pcbi.1006222> PMID: 29879103
70. Jin L, Shi YZ, Feng CJ, Tan YL, Tan ZJ. Modeling structure, stability, and flexibility of double-stranded RNAs in salt solutions. *Biophys J.* 2018; 115: 1403–1416. <https://doi.org/10.1016/j.bpj.2018.08.030> PMID: 30236782
71. SantaLucia J Jr., Allawi HT, Seneviratne PA. Improved nearest-neighbor parameters for predicting DNA duplex stability. *Biochemistry.* 1996; 35: 3555–3562. <https://doi.org/10.1021/bi951907q> PMID: 8639506
72. Zhang J, Lin M, Chen R, Wang W, Liang J. Discrete state model and accurate estimation of loop entropy of RNA secondary structures. *J Chem Phys.* 2008; 128: 125107. <https://doi.org/10.1063/1.2895050> PMID: 18376982
73. Zhang J, Zhang YJ, Wang W. An RNA base discrete state model toward tertiary structure prediction. *Chin Phys Lett.* 2010; 27: 118702.
74. Manning GS. The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. *Q Rev Biophys.* 1978; 11: 179–246. <https://doi.org/10.1017/s0033583500002031> PMID: 353876
75. Tan ZJ, Chen SJ. Electrostatic correlations and fluctuations for ion binding to a finite length polyelectrolyte. *J Chem Phys.* 2005; 122: 44903. <https://doi.org/10.1063/1.1842059> PMID: 15740294
76. Tan Z, Zhang W, Shi Y, Wang F. RNA folding: structure prediction, folding kinetics and ion electrostatics. *Adv Exp Med Biol.* 2015; 827: 143–183. [https://doi.org/10.1007/978-94-017-9245-5\\_11](https://doi.org/10.1007/978-94-017-9245-5_11) PMID: 25387965
77. Leonarski F, Trovato F, Tozzini V, Les A, Trylska J. Evolutionary algorithm in the optimization of a coarse-grained force field. *J Chem Theory Comput.* 2013; 9: 4874–4889. <https://doi.org/10.1021/ct4005036> PMID: 26583407
78. Kirkpatrick S, Gelatt CD, Vecchi MP. Optimization by simulated annealing. *Science.* 1983; 220: 671–680. <https://doi.org/10.1126/science.220.4598.671> PMID: 17813860
79. Li J, Zhang J, Wang J, Li W, Wang W. Structure prediction of RNA loops with a probabilistic approach. *PLoS Comput Biol.* 2016; 12: e1005032. <https://doi.org/10.1371/journal.pcbi.1005032> PMID: 27494763
80. Madras N, Sokal AD. The pivot algorithm: A highly efficient Monte Carlo method for the self-avoiding walk. *J Stat Phys.* 1988; 50: 109–186.
81. Mavrantzas VG. Using Monte Carlo to simulate complex polymer systems: recent progress and outlook. *Front Phys.* 2021; 9: 661367.
82. Ouldridge TE, Louis AA, Doye JP. Extracting bulk properties of self-assembling systems from small simulations. *J Phys Condens Matter.* 2010; 22: 104102. <https://doi.org/10.1088/0953-8984/22/10/104102> PMID: 21389436
83. Privalov PL, Crane-Robinson C. Translational entropy and DNA duplex stability. *Biophys J.* 2018; 114: 15–20. <https://doi.org/10.1016/j.bpj.2017.11.003> PMID: 29320682
84. Chen Q, Luo W, Veach RA, Hickman AB, Wilson MH, Dyda F. Structural basis of seamless excision and specific targeting by piggyBac transposase. *Nat Commun.* 2020; 11: 3446. <https://doi.org/10.1038/s41467-020-17128-1> PMID: 32651359

85. Wu B, Girard F, van Buuren B, Schleucher J, Tessari M, Wijmenga S. Global structure of a DNA three-way junction by solution NMR: towards prediction of 3H fold. *Nucleic Acids Res.* 2004; 32: 3228–3239. <https://doi.org/10.1093/nar/gkh645> PMID: 15199171
86. LeBlanc DA, Morden KM. Thermodynamic characterization of deoxyribooligonucleotide duplexes containing bulges. *Biochemistry.* 1991; 30: 4042–4047. <https://doi.org/10.1021/bi00230a031> PMID: 2018770
87. Williams AP, Longfellow CE, Freier SM, Kierzek R, Turner DH. Laser temperature-jump, spectroscopic, and thermodynamic study of salt effects on duplex formation by dGCATGC. *Biochemistry.* 1989; 28: 4283–4291. <https://doi.org/10.1021/bi00436a025> PMID: 2765487
88. Peyret N, Seneviratne PA, Allawi HT, SantaLucia J Jr. Nearest-neighbor thermodynamics and NMR of DNA sequences with internal A.A, C.C, G.G, and T.T mismatches. *Biochemistry.* 1999; 38: 3468–3477. <https://doi.org/10.1021/bi9825091> PMID: 10090733
89. Tran T, Cannon B. Differential effects of strand asymmetry on the energetics and structural flexibility of DNA internal loops. *Biochemistry.* 2017; 56: 6448–6459. <https://doi.org/10.1021/acs.biochem.7b00930> PMID: 29141138
90. Reiling C, Khutsishvili I, Huang K, Marky LA. Loop contributions to the folding thermodynamics of DNA straight hairpin loops and pseudoknots. *J Phys Chem B.* 2015; 119: 1939–1946. <https://doi.org/10.1021/jp5116417> PMID: 25584896
91. Rentzeperis D, Shikiya R, Maiti S, Ho J and Marky LA. Folding of intramolecular DNA hairpin loops: enthalpy-entropy compensations and hydration contributions and hydration contributions. *J Phys Chem B.* 2002; 106: 9945–9950.
92. Rentzeperis D, Kharakoz DP, Marky LA. Coupling of sequential transitions in a DNA double hairpin: energetics, ion binding, and hydration. *Biochemistry.* 1991; 30: 6276–6283. <https://doi.org/10.1021/bi00239a028> PMID: 2059634
93. Hilbers CW, Haasnoot CA, de Bruin SH, Joordens JJ, van der Marel GA, van Boom JH. Hairpin formation in synthetic oligonucleotides. *Biochimie.* 1985; 67: 685–695. [https://doi.org/10.1016/s0300-9084\(85\)80156-5](https://doi.org/10.1016/s0300-9084(85)80156-5) PMID: 4084598
94. Germann MW, Kalisch BW, Lundberg P, Vogel HJ, van de Sande JH. Perturbation of DNA hairpins containing the EcoRI recognition site by hairpin loops of varying size and composition: physical (NMR and UV) and enzymatic (EcoRI) studies. *Nucleic Acids Res.* 1990; 18: 1489–1498. <https://doi.org/10.1093/nar/18.6.1489> PMID: 2326190
95. Goddard NL, Bonnet G, Krichevsky O, Libchaber A. Sequence dependent rigidity of single stranded DNA. *Phys Rev Lett.* 2000; 85: 2400–2403. <https://doi.org/10.1103/PhysRevLett.85.2400> PMID: 10978020
96. Kuznetsov SV, Ren CC, Woodson SA, Ansari A. Loop dependence of the stability and dynamics of nucleic acid hairpins. *Nucleic Acids Res.* 2008; 36: 1098–1112. <https://doi.org/10.1093/nar/gkm1083> PMID: 18096625
97. Owczarzy R, You Y, Moreira BG, Manthey JA, Huang L, Behlke MA, et al. Effects of sodium ions on DNA duplex oligomers: improved predictions of melting temperatures. *Biochemistry.* 2004; 43: 3537–3554. <https://doi.org/10.1021/bi034621r> PMID: 15035624
98. Stefaniak F, Bujnicki JM. AnnapuRNA: A scoring function for predicting RNA-small molecule binding poses. *PLoS Comput Biol.* 2021; 17: e1008309. <https://doi.org/10.1371/journal.pcbi.1008309> PMID: 33524009
99. Wang J, Zhao Y, Zhu C, Xiao Y. 3dRNAscore: a distance and torsion angle dependent evaluation function of 3D RNA structures. *Nucleic Acids Res.* 2015; 43: e63. <https://doi.org/10.1093/nar/gkv141> PMID: 25712091
100. Xiong P, Wu R, Zhan J, Zhou Y. Pairing a high-resolution statistical potential with a nucleobase-centric sampling algorithm for improving RNA model refinement. *Nat Commun.* 2021; 12: 2777. <https://doi.org/10.1038/s41467-021-23100-4> PMID: 33986288
101. Zhao H, Yang Y, Zhou Y. Structure-based prediction of DNA-binding proteins by structural alignment and a volume-fraction corrected DFIRE-based energy function. *Bioinformatics.* 2010; 26: 1857–1863. <https://doi.org/10.1093/bioinformatics/btq295> PMID: 20525822
102. Tan YL, Wang X, Shi YZ, Zhang W, Tan ZJ. rsRNASP: A residue-separation-based statistical potential for RNA 3D structure evaluation. *Biophys J.* 2022; 121: 142–156. <https://doi.org/10.1016/j.bpj.2021.11.016> PMID: 34798137
103. Li J, Zhu W, Wang J, Li W, Gong S, Zhang J, et al. RNA3DCNN: Local and global quality assessments of RNA 3D structures using 3D deep convolutional neural networks. *PLoS Comput Biol.* 2018; 14: e1006514. <https://doi.org/10.1371/journal.pcbi.1006514> PMID: 30481171