# Atomistic Molecular Dynamics Simulations of DNA Minicircle Topoisomers: A Practical Guide to Setup, Performance and Analysis

Thana Sutthibutpong1, Agnes Noy2 and Sarah Harris2,3

1. King Mongkut University of Technology Thonburi, 126 Pracha Uthit Road, Thun Khru, Bangkok, Thailand, 10140.

2. School of Physics and Astronomy, University of Leeds, Leeds, UK, LS2 9JT.

3. Astbury Centre for Structural and Molecular Biology, University of Leeds, Leeds, UK, LS2 9JT.

## Summary

While DNA supercoiling is ubiquitous *in vivo*, the structure of supercoiled DNA is more challenging to study experimentally than simple linear sequences because the DNA must have a closed topology in order to sustain superhelical stress. DNA minicircles, which are closed circular double stranded DNA sequences typically containing between 60 and 500 base pairs, have proven to be useful biochemical tools for the study of the mechanics of supercoiled DNA. We present detailed protocols for constructing models of DNA minicircles *in silico*, for performing atomistic molecular dynamics (MD) simulations of supercoiled minicircle DNA and for analysing the results of the calculations. These simulations are computationally challenging due to the large system sizes. However, improvements in parallel computing software and hardware promise access to improve conformational sampling and simulation timescales. Given the concurrent improvements in the resolution of experimental techniques such as atomic force microscopy (AFM) and cryo-electron microscopy, the study of DNA minicircles will provide a more complete understanding of both the structure and the mechanics of supercoiled DNA.

**Key words:** Atomistic molecular dynamics, DNA supercoiling

## 1. Introduction

DNA supercoiling is ubiquitous *in vivo*, and is implicated in both genome organisation and gene regulation ***(1)***. DNA packaging in both prokaryotic and eukaryotic chromosomes can be investigated by HiC experiments, which detect spatial proximity between pairs of genomic loci through chemical cross-linking ***(2)***, and alternatively, by optical techniques such as fluorescence *in situ* hybridization (FISH) which visualises the spatial relationship of a few chosen genomic sites ***(3)***. Simple polymer models including the supercoiling of DNA suggest that topology is a crucial factor to reproduce the experimental spatial contact maps ***(4)***. Moreover, supercoiling directly induces strand separation ***(******5, 6)*** for RNA polymerase binding ***(******7)*** and altered the DNA geometry, which influences the regulatory protein binding and makes DNA able to direct its own metabolism by altering the level of supercoiling ***(******8)***.

**DNA minicircles as Model Systems:** While X-ray crystallography and NMR studies have provided atomistic resolution structural information for short (around 30 base pairs) linear DNA fragments, for protein-DNA complexes ***(9, 10)*** and even for nucleosomes ***(11)***, it has not been possible to visualise supercoiled DNA minicircles in atomistic detail because they are too structurally disordered to crystallise, they can contain kinks and defects in their structures due to the superhelical stress ***(12, 13)*** and they are too large for structural NMR. DNA minicircles have proven to be a more tractable model system for probing complex DNA topologies, because their smaller size significantly reduces the number of conformations. DNA minicircles have been studied experimentally using low resolution structural techniques such as gel electrophoresis ***(14)***, atomic force microscopy (AFM) ***(15, 16)***, cryo-electron microscopy (cryo-EM) ***(13, 17, 18)*** and cryo-electron tomography (cryo-ET) ***(19)***. While these techniques have provided information about the global shape of the DNA and how this changes in response to supercoiling, so far a fully atomistic description has only possible using computer simulation. Even the smallest DNA minicircles (around 60 base pairs ***(12)***) produce a simulation box that is large in comparison with the linear sequences conventionally studied by the DNA simulation community (which are typically between 10 and 20 base pairs in length ***(20)***), these calculations are nevertheless feasible, albeit with shorter simulation trajectories and poorer sampling. The fortuitous fact that minicircles are just large enough for low resolution structural studies while being within the reach of atomistic simulation has enabled us to employ a combination of these complementary biophysical techniques to gain insight into the atomistic structure of supercoiled DNA ***(19, 21)***.

**Overview of Simulation Protocols for DNA Minicircles:** We have developed a bespoke series of protocols for simulating supercoiled DNA minicircles, that are summarised schematically in **Fig. 1**. Each numbered step within the protocols indicated in the flow diagram is described in detail in the Methods (section 3).

Atomistic simulations of supercoiled DNA minicircles start from a simple planar circular structure constructed from the specific DNA sequence required (*step 1.*). The chosen linking number is imposed by adjusting the DNA helical twist within this planar circle (see **Fig. 2a**). This then determines the superhelical density of the closed circular loop. After creating parameters and other essential input files (*steps 2-4.*), we use a frictionless implicit solvent Generalised Born Surface Approximation (GB/SA) model, which attempts to reproduce the electrostatic screening effect of the high dielectric solvent environment (*steps 5-7.*). Our estimates suggest that this speeds up conformational changes into compacted, writhed structures by at least a factor of 10 ***(19)***. However, the initial planar DNA conformation is highly artificial, and may place unphysical levels of conformational stress upon the duplex. Moreover, while in principle GB/SA simulations provide the same conformational ensemble for the DNA but in an accelerated timeframe, in practice, the neglect of discrete water molecules and solvent ions has a non- negligible effect on the structure of the DNA. Firstly, the duplex is marginally less stable in the implicitly solvated calculations, therefore, any kinks or denaturation we observe in the absence of explicit water cannot be ascribed to the levels of superhelical stress. We have also observed that the simple Debye-Huckel electrostatic screening approximation that is used to mimic the effect of salt is not as efficient at screening the repulsion within the DNA backbone as discrete counterions. Consequently, in explicit water writhed structures have a tendency to be more compact, particularly in the presence of divalent cations such as calcium. Moreover, we have occasionally observed unphysically narrow minor grooves with the implicit solvent model. To deal with these simulation artefacts, we gently equilibrate the DNA as it relaxes from the planar starting structure into a writhed configuration to ensure that the minicircles are not irreversibly distorted. The integrity of the duplex is maintained using restraints placed on the hydrogen bonding interactions between complementary base pairs (this uses the facility for imposing NMR distance restraints within AMBER) whenever we are using the GB/SA implicit solvent approximation, prior to solvation in explicit water.

During the implicitly solvated calculations we commonly observe such large oscillations in the writhe (approximately ±0.3) that it is not possible to define a unique configuration. Instead, we use a conformational clustering analysis algorithm to select representative structures (*steps 8-9.*), and subject as many configurations as is computationally feasible to explicit solvation to obtain a more reliable atomistic description (see **Fig. 2b**); typically we select three configurations, one from each of the most populated clusters ***(19)***. These are then explicitly solvated within the GROMACS program (*steps 10-18.*). A list of essential software tools, important source files and a careful simulation protocols are provided in the Materials and Methods sections.

For the smaller DNA minicircles, the writhing transition is suppressed by the large bending energy necessary to maintain the circular form ***(22, 23)***. Thus, for the simulations of DNA minicircles with the length < 150 bp (the DNA persistence length), we skip *steps 5-9*. (GB/SA MD runs and clustering analysis) and start preparing the system for explicit water and ions solvation using GROMACS from the starting structure of the planar minicircle.

**Computer Hardware:** Simulations of DNA minicircles require specialist high performance computing facilities (HPC), such as the UK supercomputer ARCHER ***(24)*** or local HPC such as the University of Leeds supercomputers ARC1 and ARC2, where available. Running simulations in parallel over multiple CPU cores can effectively reduce the computing time. Within MD codes such as AMBER, GROMACS and NAMD, parallelisation is typically achieved through “domain decomposition”. As a simulation progresses, partitioned spatial domains within the simulation box are assigned to each of the processors, and the Newtonian physics associated with the motion of the particles within each partitioned domain will be performed by its assigned processor ***(25)*,** enabling multiple calculations to be performed concurrently. **Table 1** compares the performance of parallel MD simulation processing performed by the University of Leeds ARC1/2 local HPC service and the UK ARCHER supercomputers on the implicitly and explicitly solvated DNA minicircles performed by our group.

**Computer Software and MD forcefields:** To perform simulations at the fully atomistic level, the biomolecular simulation field has provided researchers with a wealth of tools for simulations and trajectory analysis for proteins and nucleic acids, which are fast and user friendly ***(24)***. among the most popular open-source MD packages are AMBER ***(26)***, GROMACS ***(25)*** and NAMD ***(27)***. A choice of forcefields is also available; most commonly users must choose between the AMBER suite of forcefields ***(28–30)*** and those provided by the CHARMM community ***(31)***. The parmbsc0 AMBER forcefield has become the default parameterisation due to its capacity to produce stable MD simulations of DNA duplexes at the microsecond time scale ***(32)*** and to describe a variety of non-canonical DNA structures; for minicircles using a subsequent refinement of the gamma torsion parameter from DNA backbone ***(29)*** allows a better description of severe distortions of the DNA duplex caused by strong mechanical stress. In parallel, CHARMM also provides good descriptions of DNA structure and dynamics ***(33)*** and has performed better for nanotechnology applications, such as, DNA origami ***(34)***. Although further testing is required in simulating circular DNA, the most convenient parameter combination we have found so far is the AMBER parm99+bsc0+OL4 forcefield sets..

**Comparison of DNA Minicircle Simulations with Experiments:** While MD simulations are invaluable for the study of DNA minicircles because they are the only method capable of providing atomistically detailed information, it is always desirable to validate computer models against experiment wherever possible. For DNA minicircles, this can be achieved by comparing the results of the simulations with low resolution structural studies and biochemical analysis.

One opportunity for validating the models arises because extreme bending and torsional stress can drive the DNA structures beyond their elastic regime, resulting in the formation of non-canonical DNA structures such as kinks and denaturation bubbles ***(5, 35)*** which can be detected biochemically using nuclease enzymes such as BAL-31 and S1 ***(12)*** which digest single stranded DNA. DNA defects were firstly observed in minicircle simulations at the atomistic level by Lankas *et al* ***(36)***, who reported that bending and torsional stress can give rise to type I kinks, in which the base stacking interactions are disrupted but all complementary hydrogen bonding interactions remain intact, and type II kinks, in which hydrogen bonding is disrupted. Subsequent simulations that explored higher superhelical densities also observed denaturation bubbles in which longer stretched of DNA melted into single stranded regions ***(21, 22)***. Examples of defects in DNA due to bending and superhelical stress are shown in **Fig. 2c**.

Direct visualisation of minicircles in the size range of 100-400 base pairs has also been achieved with cryo-EM ***(13)*** and AFM ***(37)***, which has enabled the comparison of static experimental minicircles structures from to be compared with simulations. A combination of gel electrophoresis, MD simulations and cryo-ET have also been used to investigate the conformational diversity of supercoiled minicircle DNA ***(19)***. Cryo-ET can provide 3D traces of the DNA minicircle helical axis for different topoisomerases, which can be compared with MD simulation trajectories. The cryo-ET shows that there is a diverse conformational ensemble for the minicircles, even for a single topoisomer. Similarly, MD simulations performed in implicit solvent show that the writhe of the minicircles is subject to large fluctutations, however, when this more approximate solvation model is used, the conformational diversity of the DNA is restricted by the hydrogen bond restraints that are applied to maintain the stability of the duplex. Subsequent simulations for discrete conformers selected from this ensemble showed that the DNA becomes more compact in explicit water due to the inclusion of discrete coaunterions that facilitates the closer approach of the two strands at the crossing point. Also, the DNA forms kink and bubble defects at high superhelical densities that increase the local flexibility of the DNA. However, the global conformational flexibility of the DNA is severely curtailed by the huge viscosity of the surrounding solvent, which effectively “locks” the DNA into a given writhed conformation over the timescale accessible to explicit MD (~50ns).

**Analysis Tools for DNA Minicircle Simulations:** The simulation community has access to a wide choice of sophisticated software for processing the velocity and coordinate data generated by MD simulations in order to both: 1) validate the simulation results by comparing the analysed data with experiments and 2) obtain structural and dynamical data inaccessible by experiment alone.

Many of these tools are available as part of the MD simulation codes. The radius of gyration of the minicircles is a global structural parameter that quantifies the compactness of supercoiled DNA, and can be extracted using the PTRAJ module available within the AMBERTOOLS software (see section 3.2). **Fig. 3a** shows three types of supercoiled DNA structures. The ‘open’ circular conformation has the largest value of radius of gyration and is the least compact structure compared to the other two. Structural disruptions and defects within the minicircles that arise due to torsional or bending stress can also be detected and quantified by measuring the interatomic distances between complementary hydrogen bonding using PTRAJ, or using visualisation software such as VMD ***(38)***. VMD can also be used to convert GROMACS trajectories into AMBER trajectories, which are compatible with the conformational analysis performed by PTRAJ.

However, other structural quantities that are specific to DNA require more specialised software. The DNA helical parameters that characterise the relative positions of the base pairs and base steps, such as twist, roll, tilt etc. can be calculated using either the CURVES+ program ***(39)*** or X3DNA ***(40)***. In addition to the radius of gyration, the overall global shape and compactness of the DNA is determined by the writhe, which quantifies the number of crossings formed by the double helix in three dimensions. Quantifying the writhe is non-trivial because the number of times that the two DNA strands cross is extremely sensitive to the precise manner in which you view the structure, as shown in **Fig. 3b**. Mathematically, the writhe is defined by the Gauss integral calculated over the central helical axis (see **Fig. 3c**) ***(41)***. However, defining an appropriate helical axis for a fully atomistic DNA structure is also non-trivial, because small local deviations from an ideal B-form helix can introduce artefacts into the writhe calculation. To eliminate these difficulties, we have developed a python script ‘WrLINE.py’, which performs a running average over each DNA helical turns ***(41)***, and is available online from the CCPForge website.

**Future Developments and Perspectives:** Improvements in the description of DNA minicircles at the atomistic level depend on two principal factors: improvements in the forcefields used to describe the DNA, and improvements in the simulation timescales and conformational sampling that can be achieved. In particular, the BSC0 forcefield is known to underestimate the twist of the DNA by approximately two degrees, although this has been partly corrected by the parmOL4 forcefield correction in which the average twist is increased from 33.2° bp-1 to 33.5° bp-1. This issue will likely be the main focus of future tests and developments for simulations of closed DNA topologies, since any discrepancy in the relaxed twist makes it difficult to make a direct comparison between the behaviours of simulated and experimental circles for a given superhelical density. However, it is not straightforward to obtain a precise comparison between twist at the atomistic level, and the global twist of the DNA (which is well established to be around 34°) because the MD forcefield has been parameterised through the interactions between atoms, not directly on the DNA helical twist itself, and DNA twist is highly sequence dependent. A consequence of this is that different sequences of the same size are likely to have significantly different superhelical densities, especially in the limit of small DNA loops.

Larger simulation timescales and improved conformational sampling will be achieved as computing technology continues to improve. For example, HPC resources continuously grow in terms of the number of processors available, which will, for example, allow for more conformations of writhed minicircle structures to be performed in explicit solvent. New computer technologies will also make a contribution to improving DNA minicircle simulations. Graphics processing units (GPUs) have already been adapted for biomolecular simulations, but as yet have not been applied to minicircles due to the limitations in systems sizes subjected to the memory constraints. GPUs have recently proven to be more cost effective than conventional computer clusters with the same number of cores ***(42)***, but require a re-engineering of the software. A GB/SA implicit solvent simulation routine has been created for GPU, which can produce a 1 μs MD trajectory for ubiquitin (containing 1,231 atoms) within ~5 days ***(43)***. Moreover, the recent development of a new GB/SA model that implements a charge hydration asymmetry (CHA) term has also improved the accuracy of the GB/SA model ***(44, 45)***. The corrected water charge distribution may well eliminate the minor groove collapsing artefact seen in the GB/SA simulations of supercoiled DNA and may facilitate robust CHA-GB/SA simulations of supercoiled minicircles on GPU with a significantly reduced cost.

Another source of accelerated biomolecular simulations is the ANTON supercomputer, which is a special-purpose computer cluster built for molecular dynamics simulations ***(46)*** which have provided MD trajectories as long as 1.119 ms within 164 days (6.82 μs/day) using 128 nodes for the explicitly solvated small protein Fip35 WW domain (containing 10,000 atoms) ***(47)***. While the first incarnation of ANTON was limited to simulations containing a maximum of 200,000 atoms, significantly less than is required for a DNA minicircle, the more recent ANTON2 machine is able to perform an MD simulation of a system containing 2,000,000 atoms and is approximately 10 times faster than the original ANTON ***(48)***, which would in principle be able to simulation a DNA minicircle containing hundreds of base pairs over multiple microsecond timescales.

Atomistically detailed MD simulations of minicircles could assist in developing them as gene therapy vectors ***(49)***. These DNA “minivectors” have already been shown to efficiently transfect lymphoma cells and have proven to be more resistant to hydrodynamic shear stress during the gene therapy delivery ***(50)*** than larger plasmids. Moreover, DNA minicircles of the same size (200-400 bp) have been detected *in vivo* where they have been excised from the chromosomes ***(51)***. Understanding the dynamic properties of DNA minicircles would help to elucidate the occurrence and sequence specificity of chromosomal microdeletion processes.

## 2. Materials

### 2.1) MD Simulation, Analysis and Visualisation Software

1. *AMBER:* The molecular dynamics simulation software package, along with the forcefield ff99+bsc0+OL4 are the main tools we have used in the GB/SA implicit solvent simulation ***(******52)***. In an AMBER simulation, the module SANDER reads in a series of input parameter, topology and coordinate files to output a \*.mdcrd AMBER trajectory and a restart coordinate files ***(******53)***.
2. *GROMACS:* This MD software package currently has the reputation of being the most efficient at generating explicitly solvated simulations trajectories. Fast MD is extremely important for DNA minicircles, given that the simulation cell can contain more than 1,000,000 atoms. A similar forcefield (e.g. AMBER) as for the implicitly solvated calculations is used to maintain consistency between the two types of simulations. A perl script ‘ambgmx.pl’ is used to create GROMACS topology and coordinate files for AMBER forcefields ***(******25)***.
3. *Specialised forcefield and ions parameters:* 1) Circular structures require a slight modification to the AMBER libraries (we call this parameter set ‘ff99circ’) in which the end residues are omitted to prevent the xleap module adding them automatically 2) DNA simulations require the ‘parmbsc0’ modification for nucleic acid backbone dihedrals α and γ ***(******28)***, 3) we also use the ‘OL4’ modification for DNA backbone dihedral χ ***(29)*** and 4) the ion modification for Na and Cl counterions of Smith and Dang ***(******54)***.
4. *Visual Molecular Dynamics (VMD)* ***(******38)****:* This visualisation software is fast and efficient, and is capable of reading molecular dynamics trajectory files of any format (e.g. NAMD, AMBER, GROMACS or etc.) and of rendering 3D images for dissemination in papers and in seminar presentations. For more complex tasks, ‘\*.tcl’ scripts can be written to control the program via the ‘tk console’. We use a script that converts a GROMACS compressed trajectories into an AMBER trajectory that is compatible with PTRAJ for analysis of the conformations, while other researchers may prefer using VMD directly, or GROMACS analysis tools.

### 2.2) AMBERTOOLS Modules

1. *NAB (Nucleic Acid Builder):*This is a C-based high level programming language, which performs specific operations on the starting coordinates of nucleic acid structures. NAB is capable of importing standard reference frames of coordinates of nucleotides from an AMBER library, performing coordinate transformations and creating a PDB file containing all the coordinates atomic types for a starting structure ***(******53)***. We use NAB to build minicircles; but knots and more complex topologies are also possible.
2. *LEAP:*This module generates topology files containing the interatomic connectivities and the associated forcefield parameters (e.g. stretching, bending and dihedral angles and non-bonded interactions) from a PDB file created by NAB.
3. *PTRAJ:* This module is used for trajectory processing such as concatenation of MD trajectories and removing unwanted atoms or residues from the trajectories (e.g. stripping out the water molecules). This AMBER module also includes commands to measure distances between atoms or residues (e.g. for calculating distances between atoms that are engaged in complementary hydrogen bonds between base pairs), bending and dihedral angles, root mean square deviations between two structures and the radius of gyration.

### 2.3) Source Files

**Table 2** shows the list of input files used to prepare the start-up files for both the implicit and explicit solvent simulation protocols and the post-simulation analysis. To use these, AMBER, AMBERTOOLS, GROMACS, python (with NUMPY library) and perl software packages need to be installed.

## 3. Methods

### 3.1) Molecular Dynamics Simulations

1. *Build a circular DNA starting structure:* Run the NAB script ‘circ.nab’ to perform these operations: 1) Determine the radius of the circle and the average twist/basepair from the number of base pairs, the ‘rise’ parameter value (~3.38 Å) and the linking number. 2) Each base pair coordinate is imported and placed its centre at the circle circumference, normal to the tangent. 3) A coordinate transformation is performed to twist the base pair by the specified average twist angle per base pair relative to its predecessor. 4) The whole circle is rotated about its centre to create a space for the next base pair. Repeating this process will create a planar circular DNA with a uniform twist angle used at the MD starting structure, for which a PDB file will be created. (*see* **Note 1**)
2. *Creating topology and coordinate files:* Run the prepared LEAP script to perform the following operations: 1) load the forcefield parameter ‘ff99circ’ for circular DNA, all the preparation (all\_nuc94bsc0\_chiOL4.in ***(******29)***) and forcefield modification files for backbone dihedrals (frcmod.parmbsc0 ***(******28)*** and frcmod.OL4.chi ***(******29)***), ions (dangion.dat ***(******54)***) and the modified DNA library (DNA\_CI.lib) containing the updated version of DNA partial charges compatible with the ‘bsc0’ parameters ***(******28)***. 2) Map some of the atom names and types in the PDB files created by NAB to be compatible with LEAP by using the ‘addPdbAtomMap’ and ‘addAtomTypes’ LEAP commands (*see* **Note 2**). 3) Load the PDB file of the circular DNA created by NAB. 4) Manually create two covalent bonds to connect two backbone strands of the DNA ends by using the ‘bond’ LEAP command. 5) Use ‘saveamberparm’ LEAP command to export a .prmtop topology file and an .inpcrd coordinate file for an implicitly solvated simulation. (*see* **Notes 3,4**)
3. *Prepare AMBER input files for GB/SA implicit solvent simulations:* 1) *two-stage minimisation* ‘min1.in’ with all-atom coordinate restraints (ntr=1) and ‘min2.in’ without coordinate restraints. 2) *three-stage equilibration with all atom coordinate restraints* ‘md1.in’: system is heated from 100 K to the constant 300 K, ‘md2.in’ and ‘md3.in’: system is under the reduced restraint weight (*see* **Table 2**). 3) *productive MD run* ‘md4.in’ the all-atom restraints have been removed and the system is restrained through its complementary base pair hydrogen bonds using the NMR distance restraints facility in SANDER (nmropt=1, pencut=-0.001). (*see* **Note 5**)
4. *Prepare an AMBER NMR distance restraints file:* To prepare the NMR restraint file, run the python script ‘genrst.py’ to read the DNA sequence and to generate an input file for the AMBER ‘makeDIST\_RST’ command. This will create another file named ‘RST’ containing all the restraint information for the productive GB/SA MD run. (*see* **Note 6**)
5. *AMBER GB/SA implicit solvent minimisation and equilibration:* Run the first minimisation stage from the prepared AMBER topology, starting coordinate and ‘min1.in’ AMBER input files using the SANDER module. The final structure of ‘min1’ will be the starting structure for ‘min2’, for which the final structure will be the starting structure for ‘md1’. This procedure will be repeated so that the final structure of ‘md3’ equilibration stage will be the starting structure for the productive MD run.
6. *AMBER GB/SA implicit solvent productive run:* From the final structure of ‘md3’, execute the ‘md4’ productive MD run in the GB/SA implicit solvent with all the hydrogen bonds restrained. If the simulation finishes before a sufficiently long trajectory has been obtained, it can be extended from the final structure using the AMBER restart and topology files. (*see* **Note 7**)
7. *Cut and merge the trajectories of writhed DNA structures:* Use PTRAJ to concatenate the trajectories. Discard the first few nanoseconds (maybe 2-5 ns) of all the replicas, keeping only the snapshots where the writhe values become stable. If convenient, PTRAJ allows the user to merge all the replicas to create a single file containing the ensemble of writhed DNA structures of a given topoisomer for subsequent analysis.
8. *Clustering analysis and selection for the supercoiled DNA conformation to be explicitly solvated:* In PTRAJ, use the command ‘cluster’ with the option ‘averagelinkage’ to divide all the MD snapshots for each topoisomer into clusters (we typically generate around six clusters). Visualise the highly populated conformational clusters in VMD, and select a representative structure. Ensure that the chosen structure is free from structural disruptions, paying particular attention to the base stacking interactions (the hydrogen bond restraints prevent the formation of single stranded regions, but do not maintain stacking).
9. *Vacuum minimisation:* A representative structure of supercoiled DNA from a GB/SA simulation may contain regions where the minor groove is artificially narrow, especially at the plectoneme apices. This is an artefact of the approximate implicit solvent model, and therefore should be removed prior to adding counterions and water molecules. This minor groove compaction is relieved by a short minimisation run in vacuum (igb=0), as the electrostatic repulsion between backbone phosphate groups widens the minor groove.
10. *File conversion from AMBER to GROMACS:* the GROMACS topology (.top) and coordinate (.gro) files can be created Using the *perl* script ‘amb2gmx.pl’ (which is available from the GROMACS website) along with AMBER topology and coordinate files of the starting structure (without solvent). (*see* **Note 8,9**)
11. *Prepare the GROMACS topology files (\*.top) from modular files (.itp):* 1) ‘ffbsc0.itp’ specifies the type of forcefield to be chosen (e.g. AMBER, CHARMM or GROMOS, we use AMBER ***(******28)***), 2) ‘molecule.itp’ contains all the forcefield parameters of the molecule to be studied, 3) ‘tip3p.itp’ contains the topology of a TIP3P water molecule, 4) ‘ions.itp’ contains the topology of the chosen model of ions (Dang and Smith in our case ***(54)***) and 5) ‘posre.itp’ contains force constants for coordinate restraints used during the equilibration. The \*.itp files will be specify as necessary inside the \*.top files to build the final topology as a structural modular parts.
12. *Explicit water solvation in GROMACS:* from the GROMACS topology and coordinate files created by ‘amb2gmx.pl’, the following commands should be executed to solvate the molecule in explicit water: 1) By using ‘editconf’, the structure is centred within a triclinic box and its principal axis is set to be parallel to the box axis (with the option ‘-princ’). 2) ‘genbox’ is used to generate a coordinate file of the molecule in a TIP3P water box ‘molecule.w.gro’ with the option to load the reference coordinates of water molecules (-cs spc216.gro) (*see* **Note 10**). 3) A run input file ‘molecule.w.tpr’ is generated from ‘molecule.w.gro’ and a topology file ‘molecule.w.top’ containing ‘ffbsc0.itp’, ‘molecule.itp’, ‘tip3p.itp’, ‘ions.itp’, and the exact number of water molecules is assigned using the ‘grompp’ command.
13. *Explicit ion solvation in GROMACS:* 1) The ‘genion’ command with the option ‘-norandom’ (*not* included in GROMACS 5) is used to introduce a number of positive monovalent Na+ counterions, which replace water molecules around the negatively charged phosphate groups. This creates a coordinate file ‘molecule.wp.gro’, in which the system is electrically neutralised. 2) A run input file ‘molecule.wp.tpr’ is generated by the ‘grompp’ command from the files ‘molecule.wp.gro’ and the topology file ‘molecule.wp.top’. This contains all the necessary .itp files and the number of water molecules and ions corresponding to ‘molecule.wp.gro’ (see step 11). 3) If a salt concentration higher than minimal is required, then additional positive and negative ions are introduced. This uses the options ‘-nn {number of negative ions}’ and ‘-np {number of positive ions}’ in the ‘genion’ command. The number of positive and negative ions required to emulate a specific salt concentration can be calculated from 0.0018×(salt concentration in molar)×(number of water molecules in the water box). The output coordinate file is named ‘molecule.wnr.gro’ 4) A new run input file ‘molecule.wnr.tpr’ is generated from ‘molecule.wnr.gro’ and the topology file ‘molecule.wnr.top’, which contains the final numbers of water molecules and ions.
14. *Prepare the .mdp input files for GROMACS explicit solvent simulations:* the parameters for the multistep equilibration protocols were designed based on the standard simulation protocols for relaxing DNA ***(******55)*** (*see* **Table 3**), but with additional steps to relieve any additional conformational stress associated with the closed topology. This system undergoes a four-stage minimisation and eight-stage equilibration prior to the production run.
15. *Prepare the topology files for GROMACS explicit solvent simulations:* The equilibration proceeds from the run input file ‘molecule.wnr.tpr’. For each relaxation stage requiring all-atom coordinate restraints, a coordinate restraint file (e.g. ‘posre.itp’) needs to be generated using the ‘genrestr’ command. This will be included in the topology file corresponding to each minimization and equilibration (e.g. ‘molecule.min1.top’ for minimisation stage 1).
16. *Run GROMACS explicit solvent simulations:* For each minimisation and equilibration stage the ‘mdrun’ command is executed, which produces the updated restart coordinate file (e.g. ‘molecule.min1.gro’ the minimisation stage 1 etc). The command is then used to ‘grompp’ create a run input file (e.g. ‘molecule.min2.tpr’), which is executed by ‘mdrun’ in the next stage. Minimisation and equilibration continues through to the end of equilibration stage 8. Then, the productive MD run starts, in which all the coordinate restraints are lifted and DNA is free to move under the explicit solvent environment. (*see* **Notes 11,12**).
17. *Processing the GROMACS MD trajectories:* After the productive MD simulation run finishes, we process the trajectory data by using ‘trjconv’ command. 1) It is often the case that it is only solute structure that is of interest. Fully solvated trajectory files are usually inconveniently large, which means they can be slow to transfer between the supercomputer and local workstations (where much of the visualisation and analysis will be performed), they can be difficult to store, and can cause problems with memory during visualisation. The water molecules and ions can be removed from the trajectory by specifying the atomic index numbers (turn on the option ‘-n’ of ‘trjconv’) (*see* **Note 13**). 2) If required, it is also possible to reduce the time frames sampled in the trajectory using the option ‘-dt’. (*see* **Note 14).** It is also possible to specify that only the solute molecules are output in the trajectories, however, as it is sometimes desirable to analyse the water molecule positions, particularly when structure defects are present, by default we save the water molecule co-ordinates and discard them only when we are sure they are not needed.
18. *Convert GROMACS trajectories back to AMBER \*.mdcrd format:* To use PTRAJ for data analysis, is it first necessary to convert the GROMACS trajectories into AMBER format. VMD is one of the software tools that is conveniently able to load most MD trajectory formats, and can interconvert between the different formats according to the preferences of the user.
19. *Repeat the process from stage 1 to run more simulations of other topoisomers or different minicircles sizes or sequences.*

### 3.2) Conformational Analysis

1. *Calculating the Radius of gyration from the implicitly solvated trajectories:* The ‘radgyr’ command implemented in the AMBERTOOLS’s PTRAJ module is able to calculate the radius of gyration from a series of MD trajectory snapshots. We use implicitly solvated trajectories to calculate this quantity, because the frictional term associated with inclusion of explicit water impedes conformational fluctuations of the minicircle, effectively freezing the DNA into a single writhed conformation. The radius of gyration calculated from implicitly solvated trajectories of different topoisomers is a useful physical parameter that can be compared with the experimentally measured mobilities of the corresponding supercoiled minicircle topoisomers within polyacrylamide gels. (*see* **Note 15**)
2. *Calculation of the DNA minicircle writhe using by the python script WrLINE:* To calculate the writhe, the script takes as input the AMBER topology and trajectory files. The user needs to specify the number of base pairs and the number of MD snapshots within the ‘WrLINE.py’ script. The output consists of a time series of writhe values (*see* **Note 16**).
3. *Create traces from MD trajectories to be compared with cryo-EM/ET or AFM:* To create a 3D trace of the minicircle shape, use PTRAJ to strip all the atoms from the MD trajectory, except the 20 C1’ carbon atoms that are approximately equally spaced from their neighbours, to represent the writhed structure (The number 20 is the arbitrary number of points used by our experimental collaborators to represent the cryo-ET computational trace). (*see* **Note 17**) While AFM and cryoEM give only 2D (not 3D) information, the trace is nevertheless useful for comparing with these experimental data.
4. *Calculate the RMSD between the explicitly solvated DNA to cryo-ET traces:* Given a set of computational traces obtained from cryo-ET density maps, it is possible to quantify the level of agreement between the experiments and the simulations by using the ‘rms’ command in PTRAJ to calculate the root mean square deviation (RMSD) values for all pairs of the discrete points that constitute the MD and cryo-ET traces. We assign the MD trace that gives the smallest RMSD value with any of the cryo-ET traces to be the ‘best representative’ atomistic structure for the supercoiled DNA.

## 4. Notes

1. To build a DNA starting structure, a script containing commands and operations on the coordinates of DNA base pairs can be written by using the NAB scripting language, which is implemented in AMBERTOOLS. This is then compiled as an executable file. Given the required number of base pairs, the radius of DNA the circle can be specified (this is calculated from the rise), a circle can be drawn from that point and the average local twist angle can be given. For example, for a 336 bp DNA, with the linking number Lk = 30 and the ‘rise’ parameter (the optimal stacking distance between basepairs) of 3.38 Å, the circle has the average twist angle between two neighbouring basepairs of (30×360°) / 336 = 32.14°.
2. Either XLEAP (the graphical version of LEAP) or TLEAP (the text-based version) can be used to create the basic start-up files topology and co-ordinate files for an AMBER MD simulation. To ensure that the atom naming conventions are compatible with the output from NAB, the ‘addPdbAtomMap’ command should be used to map the following atomic nomenclatures: {{OP1 O1P}{OP2 O2P}{H5’ H5’1}{H5’’ H5’2}{H2’ H2’1}{ H2’’ H2’2}}, and two additional ringed carbon atom types “C1” and “C2” are defined by ‘addAtomTypes {{“C1” “C” “sp2”}{“C2” “C” “sp2”}}’.
3. The coordinate file (.inpcrd in AMBER7 or later) contains all of the position and velocity information for the system necessary to restart the simulation and continue a dynamics run. The topology file (.mdcrd in AMBER7 or later) contains all the necessary physical chemical information for describing each atom and its interactions (mass, radius, partial charge and etc.) and covalent bonding (e.g. stretching, bending and dihedral force constants)
4. With the information from topology and coordinate (restart) files, a PDB file of a molecule can be reconstructed using the AMBER utility script ‘ambpdb’. This operation can be useful for visualising the molecule to quickly check the starting co-ordinates, or to check progress during the simulation (as restart files are continuously output as the MD progresses), or as a way of identifying problems if a simulation crashes.
5. In the absence of friction from the hydrodynamic interactions associated with explicit water molecules, we use either all-atom or NMR distance restraints between complementary base pairs to prevent structural disruptions within the DNA, which could be due to approximated solvent model, as this generally destabilises the DNA. All of the restraints are removed only when explicit solvent is used, which is the only time that we can observed defects involving broken hydrogen bonds. The option ‘pencut=-0.001’ is set to have a negative value in order to print out all the energy associated with the distance deviation from the NMR restraints.
6. The input file for NMR restraint generation contains eight columns for each line representing a distance restraint: ‘1 ADE N1 8 THY H3 1.70 2.10’ the first six columns identify residue numbers, residue types and atomic types of a pair of restrained atoms, and the last two columns describe the restraint boundaries.
7. To obtain replica trajectories to improve conformational sampling, simply run ‘md4’ but assigning a new set of velocities to the DNA from the restart file (irest=1, ntx=5). Running dynamics using an independent set of initial starting velocities will cause the trajectories to diverge in conformational space, even if the initial atomic co-ordinates are the same.
8. In order to explicitly solvate the 336 bp supercoiled DNA minicircles, a large water box containing around 700,000 water molecules (over 2,000,000 atoms) is required. AMBER is not currently capable of simulating systems containing more than 1,000,000 atoms. Thus, GROMACS is used to carry out these large systems. On our supercomputer resources, GROMACS is currently the most efficient MD engine for explicitly solvated calculations.
9. For smaller systems (e.g. minicircles containing around 100 base pairs), we are able to carry out the explicitly solvated MD simulations in AMBER. Additional LEAP commands are added to solvate the circular starting DNA structure in an explicit solution: 1) including a TIP3P water box by using ‘solvatebox’ command, 2) neutralising the system by a number of positive monovalent counterions (Na+, K+ or etc.) by using ‘addions’ command, 3) adding further ions to emulate a specified salt concentration.
10. For a short explicit solvent simulation (10-20 ns) of a 336 bp supercoiled minicircle that involve no major changes in DNA writhe, a triclinic or rectangular waterbox with 30-40 Å solution buffer is an optimal choice to prevent clashing of a pair of DNA segments from neighbouring periodic box.
11. In productive MD runs within GROMACS, the trajectory files with full precision produced by ‘mdrun’ are in the ‘\*.trr’ format (option ‘-o’). Alternatively, to significantly reduce the file size, the option ‘-x’ can be used to output the compressed MD trajectory in the ‘\*.xtc’ format.
12. Many supercomputing systems have their own time limit for running a job, which is typically between 12 and 48 hours. Therefore, if one needs to produce a long MD trajectory (e.g. 100ns), simulations need to be conducted by continuously resubmitting the job from the restart files produced by a previous run. When the trajectory reaches the length required (or the user runs out of patience!), all the separate trajectories can be catenated using the GROMACS command ‘trjcat’.
13. The GROMACS module ‘make\_ndx’ is a tool for obtaining the atomic index numbers of any specified groups (e.g. DNA, water, ions, non-water), corresponding to the index numbers in the \*.gro coordinate files. Indices produced by ‘make\_ndx’ serves as input files for any GROMACS commands performed on specific groups of atoms.
14. In the frequent instance when sections of the solute “jump out” of the primary periodic water box, this can spoil the visualisation and create errors in conformational analysis. Using ‘trjconv’ with the option ‘-pbc’ can repair this problem. (For further details, please consult GROMACS manual)
15. Errors can occur when comparing the radius of gyration calculated from the ensemble of implicitly solvated MD snapshots of supercoiled DNA minicircles with experimental data due to need to impose hydrogen bond restraints when using approximate solvent models. The artificial hydrogen bonding restraints prevent the DNA relieving torsional or bending stress through structural disruptions and defect formation. These errors are most severe for highly supercoiled minicircles that are under the most torsional stress. Therefore, it is most important to perform explicitly solvated calculations for topoisomers at higher superhelical densities.
16. The PTRAJ module and the NUMPY python library are required in order to run the ‘WrLINE.py’ script.
17. To pick 20 representative atoms from MD trajectories to be compared with the cryo-ET traces of 336 bp minicircles (1 point represents 16.8 basepairs), for example, C1’ atoms can be picked from the basepair 17, 34, 51, 68, 84, …, 336.

Specimen inputs files and scripts to build the DNA structures, perform the simulations and analysis described are provided as Supplementary Information in the file “Source.zip”.

## 5. References

1. A.D. Bates and A. Maxwell (2005) DNA Topology, Oxford University Press.

2. J. Dekker, M. a Marti-Renom, and L. a Mirny (2013) Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data., Nature reviews. Genetics. 14, 390–403.

3. I. Williamson, S. Berlivet, R. Eskeland, et al. (2014) Spatial genome organization: contrasting views from chromosome conformation capture and fluorescence in situ hybridization, Genes & Development. 28, 2778–2791.

4. F. Benedetti, A. Japaridze, J. Dorier, et al. (2015) Effects of physiological self-crowding of DNA on shape and biological properties of DNA molecules with various levels of supercoiling, Nucleic Acids Research. 43, 2390–2399.

5. H.R. Drew, J.R. Weeks, and A.A. Travers (1985) Negative supercoiling induces spontaneous unwinding of a bacterial promoter, EMBO Journal. 4, 1025 –1032.

6. F. Kouzine, D. Levens, and L. Baranello (2014) DNA topology and transcription., Nucleus. 5, 195–202.

7. I.K. Pemberton, G. Muskhelishvili, A.A. Travers, et al. (2002) FIS modulates the kinetics of successive interactions of RNA polymerase with the core and upstream regions of the tyrT promoter, Journal of Molecular Biology. 318, 651–663.

8. J.M. Fogg, G.L. Randall, B.M. Pettitt, et al. (2012) Bullied no more: when and how DNA shoves proteins around., Quarterly reviews of biophysics. 45, 257–299.

9. W.K. Olson, A.A. Gorin, X.-J. Lu, et al. (1998) DNA sequence-dependent deformability deduced from protein–DNA crystal complexes, Proceedings of the National Academy of Science USA. 95, 11163–11168.

10. N.M. Luscombe, S.E. Austin, H.M. Berman, et al. (2000) An overview of the structures of protein-DNA complexes., Genome biology. 1, REVIEWS001.

11. T.J. Richmond and C.A. Davey (2003) The structure of DNA in the nucleosome core, Nature. 423, 145–150.

12. Q. Du, A. Kotlyar, and A. Vologodskii (2008) Kinking the double helix by bending deformation, Nucleic Acids Research. 36, 1120–1128.

13. T.A. Lionberger, D. Demurtas, G. Witz, et al. (2011) Cooperative kinking at distant sites in mechanically stressed DNA., Nucleic acids research. 39, 9820–9832.

14. J.M. Fogg, N. Kolmakova, I. Rees, et al. (2006) Exploring writhe in supercoiled minicircle DNA., Journal of physics. Condensed matter : an Institute of Physics journal. 18, S145–S159.

15. L.S. Shlyakhtenko, V.N. Potaman, R.R. Sinden, et al. (1998) Structure and dynamics of supercoil-stabilized DNA cruciforms., Journal of molecular biology. 280, 61–72.

16. T. Schmatko, P. Muller, and M. Maaloum (2014) Surface charge effects on the 2D conformation of supercoiled DNA., Soft matter. 10, 2520–9.

17. J. Bednar, P. Furrer, A. Stasiak, et al. (1994) The Twist, Writhe and Overall Shape of Supercoiled DNA Change During Counterion-induced Transition from a Loosely to a Tightly Interwound Superhelix, Journal of molecular biology. 235, 825–847.

18. A. Amzallag, C. Vaillant, M. Jacob, et al. (2006) 3D reconstruction and comparison of shapes of DNA minicircles observed by cryo-electron microscopy., Nucleic acids research. 34, e125.

19. R.N. Irobalieva, J.M. Fogg, D.J. Catanese, et al. (2015) Structural diversity of supercoiled DNA, Submitted to Nature Communication.

20. D.L. Beveridge, T.E. Cheatham, and M. Mezei (2012) The ABCs of molecular dynamics simulations on B-DNA, circa 2012, Journal of Biosciences. 37, 379–397.

21. J.S. Mitchell, C.A. Laughton, and S.A. Harris (2011) Atomistic simulations reveal bubbles, kinks and wrinkles in supercoiled DNA, Nucleic Acids Research. 39, 3928–3938.

22. S.A. Harris, C.A. Laughton, and T.B. Liverpool (2008) Mapping the phase diagram of the writhe of DNA nanocircles using atomistic molecular dynamics simulations., Nucleic acids research. 36, 21–29.

23. T.B. Liverpool, S.A. Harris, and C.A. Laughton (2008) Supercoiling and denaturation of DNA loops, Physical Review Letters. 100, 238103.

24. A. Gray, O.G. Harlen, S.A. Harris, et al. (2015) In pursuit of an accurate spatial and temporal model of biomolecules at the atomistic level: a perspective on computer simulation, Acta Crystallographica D. D71, 162–172.

25. S. Pronk, S. Páll, R. Schulz, et al. (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit., Bioinformatics (Oxford, England). 29, 845–854.

26. D.A. Case, T.E. Cheatham, T. Darden, et al. (2005) The Amber biomolecular simulation programs., Journal of computational chemistry. 26, 1668–88.

27. J.C. Phillips, R. Braun, W. Wang, et al. (2005) Scalable molecular dynamics with NAMD, Journal of Computational Chemistry. 26, 1781–1802.

28. A. Pérez, I. Marchán, D. Svozil, et al. (2007) Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers., Biophysical journal. 92, 3817–3829.

29. M. Krepl, M. Zgarbová, P. Stadlbauer, et al. (2012) Reference simulations of noncanonical nucleic acids with different χ variants of the AMBER forcefield: quadruplex DNA, quadruplex RNA, and Z-DNA, Journal of Chemical Theory and Computation. 8, 2506–2520.

30. M. Zgarbova, F.J. Luque, S. Jir, et al. (2013) Toward Improved Description of DNA Backbone: Revisiting Epsilon and Zeta Torsion Force Field Parameters, Journal of Chemical Theory and Computation. 9, 2339–2354.

31. K. Vanommeslaeghe, E. Hatcher, C. Acharya, et al. (2009) CHARMM General Force Field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields, Journal of computational chemistry. 31, 671–690.

32. A. Pérez, F.J. Luque, and M. Orozco (2007) Dynamics of B-DNA on the microsecond time scale., Journal of the American Chemical Society. 129, 14739–45.

33. K. Hart, N. Foloppe, C.M. Baker, et al. (2012) Optimization of the CHARMM additive force field for DNA: Improved treatment of the BI/BII conformational equilibrium, Journal of Chemical Theory and Computation. 8, 348–362.

34. J. Yoo and A. Aksimentiev (2013) In situ structure and dynamics of DNA origami determined through molecular dynamics simulations., Proceedings of the National Academy of Sciences of the United States of America. 110, 20099–104.

35. F.H.C. Crick and A. Klug (1975) Kinky Helix, Nature. 255, 530–533.

36. F. Lankas, R. Lavery, and J.H. Maddocks (2006) Kinking occurs during molecular dynamics simulations of small DNA minicircles., Structure (London, England : 1993). 14, 1527–1534.

37. D. Ackermann, G. Rasched, S. Verma, et al. (2010) Assembly of dsDNA nanocircles into dimeric and oligomeric aggregates., Chemical communications (Cambridge, England). 46, 4154–4156.

38. W. Humphrey, A. Dalke, and K. Schulten (1996) VMD: Visual Molecular Dynamics, Journal of Molecular Graphic. 14, 33–38.

39. R. Lavery, M. Moakher, J.H. Maddocks, et al. (2009) Conformational analysis of nucleic acids revisited: Curves+., Nucleic acids research. 37, 5917–5929.

40. X.-J. Lu and W.K. Olson (2008) 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures., Nature protocols. 3, 1213–1227.

41. T. Sutthibutpong, S.A. Harris, and A. Noy (2015) Comparison of Molecular Contours for Measuring Writhe in Atomistic Supercoiled DNA, Journal of Chemical Theory and Computation. 150521070539002.

42. R.O. Dror, R.M. Dirks, J.P. Grossman, et al. (2012) Biomolecular Simulation: A Computational Microscope for Molecular Biology, Annual Review of Biophysics. 41, 429–452.

43. A.W. Goetz, M.J. Williamson, D. Xu, et al. (2012) Routine microsecond molecular dynamics simulations with amber - part i: Generalized born, J. Chem. Theory Comput. 8, 1542–1555.

44. A. Mukhopadhyay, A.T. Fenley, I.S. Tolokh, et al. (2012) Charge Hydration Asymmetry : The Basic Principle and How To Use It To Test and Improve Water Models Charge hydration asymmetry : the basic principle and how to use it to test and improve water models, Journal of Physical Chemistry B. 116, 9776–9783.

45. A. Mukhopadhyay, B.H. Aguilar, I.S. Tolokh, et al. (2014) Introducing Charge Hydration Asymmetry into the Generalized Born Model, Journal of Chemical Theory and Computation. 10, 1788 – 1794.

46. D.E. Shaw, M.M. Deneroff, R.O. Dror, et al. (2008) Anton, a special-purpose machine for molecular dynamics simulation, Communications of the ACM. 51, 91–97.

47. S. Piana, K. Lindorff-Larsen, and D.E. Shaw (2012) Protein folding kinetics and thermodynamics from atomistic simulation, Proceedings of the National Academy of Sciences. 109, 17845–17850.

48. D.E. Shaw, J.P. Grossman, J.A. Bank, et al. (2014) Anton 2: Raising the Bar for Performance and Programmability in a Special-Purpose Molecular Dynamics Supercomputer, SC14: International Conference for High Performance Computing, Networking, Storage and Analysis. 41–53.

49. N. Zhao, J.M. Fogg, L. Zechiedrich, et al. (2011) Transfection of shRNA-encoding Minivector DNA of a few hundred base pairs to regulate gene expression in lymphoma cells., Gene therapy. 18, 220–224.

50. D.J. Catanese, J.M. Fogg, D.E. Schrock, et al. (2012) Supercoiled Minivector DNA resists shear forces associated with gene therapy delivery., Gene therapy. 19, 94–100.

51. Y. Shibata, P. Kumar, R. Layer, et al. (2012) Extrachromosomal microDNAs and chromosomal microdeletions in normal tissues, Science (New York, N.Y.). 336, 82–86.

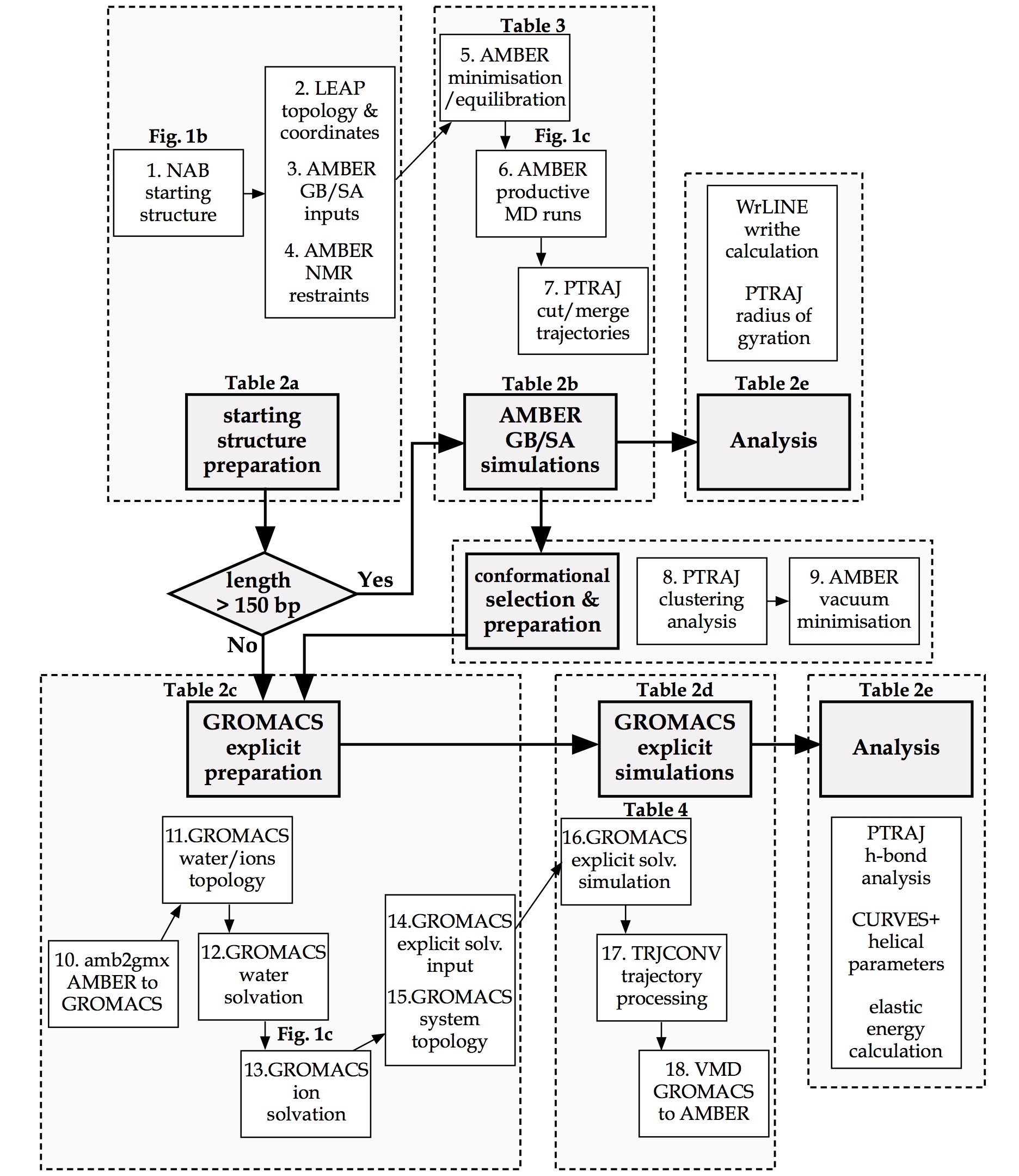
52. V. Tsui and D.A. Case (2000) Theory and Applications of the Generalized Born Solvation Model in, Biopolymers. 56, 275–291.

53. D.A. Case, T.A. Darden, T.E. Cheatham III, et al. (2010) Amber 11, University of California, San Francisco.

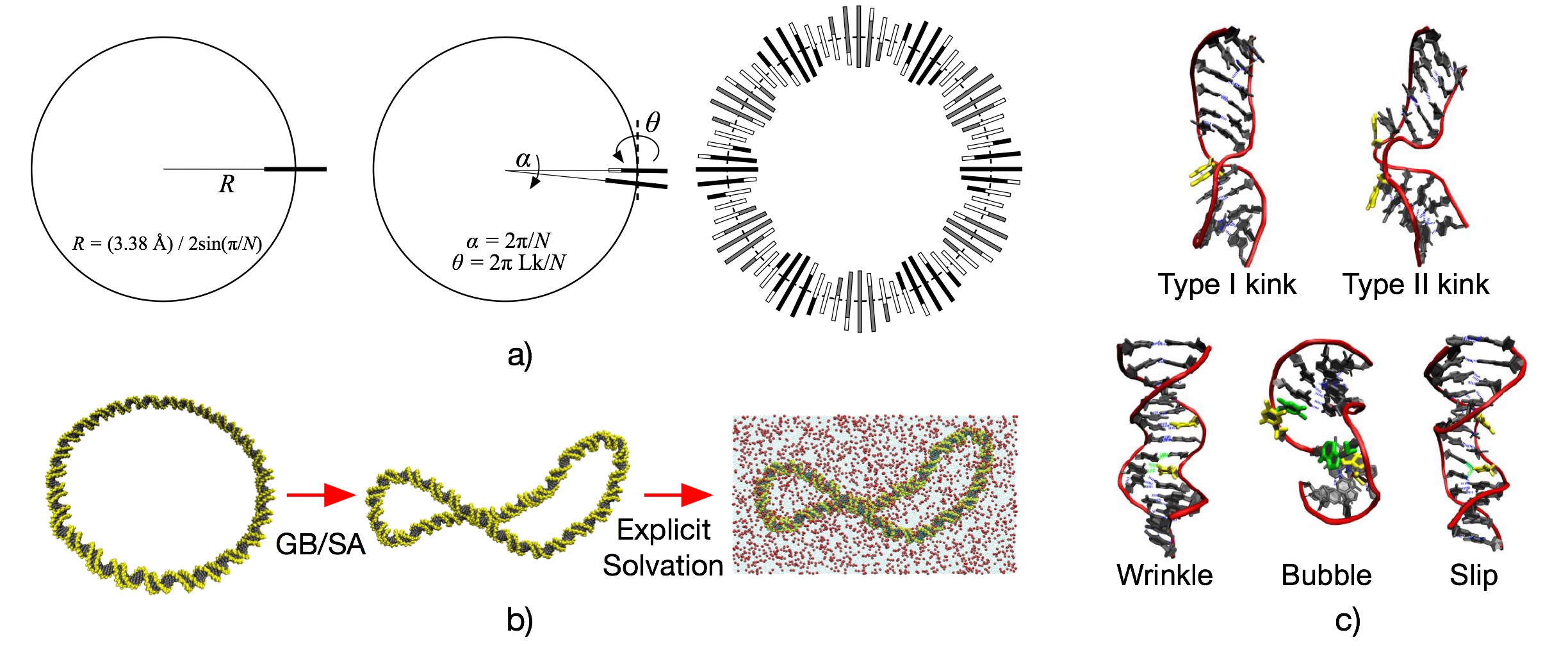
54. D.E. Smith and L.X. Dang (1994) Computer simulations of NaCl association in polarizable water, The Journal of Chemical Physics. 100, 3757.

55. G.C. Shields, C.A. Laughton, and M. Orozco (1997) Molecular Dynamics Simulations of the d(T·A·T) Triple Helix, Journal of the American Chemical Society. 119, 7463–7469.

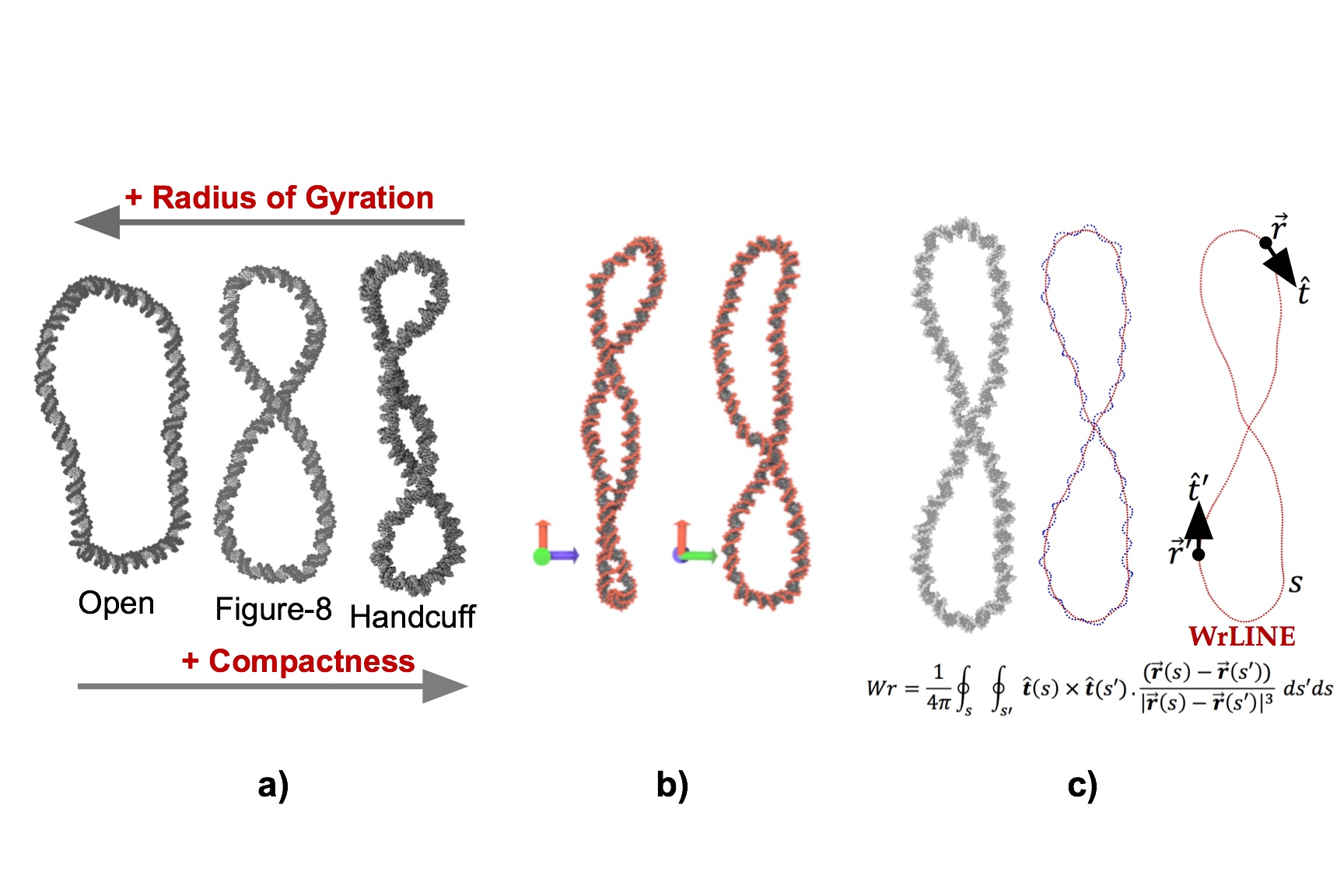
## Figures



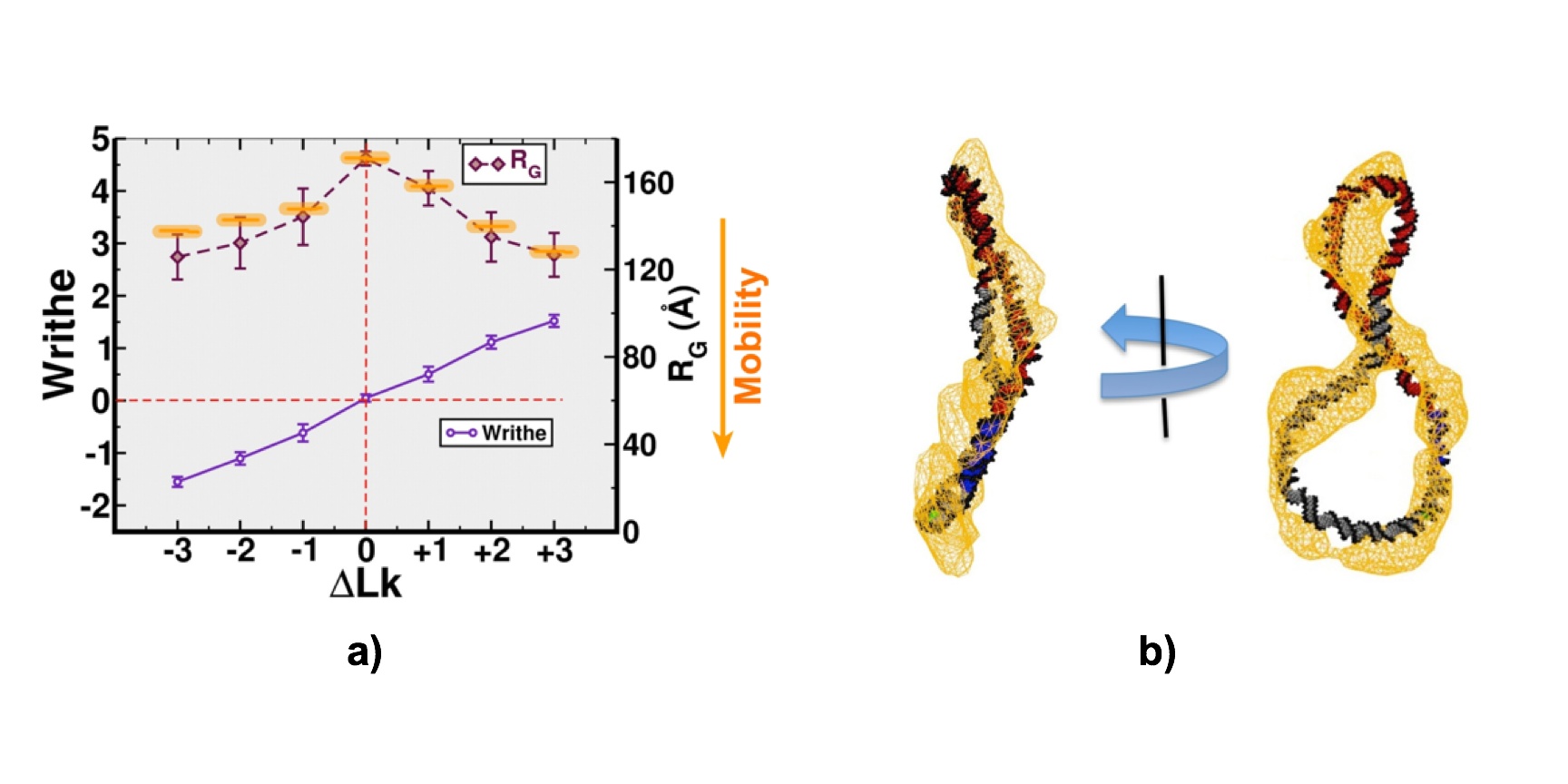
**Figure 1** Flow diagram providing an overview of DNA minicircle simulation protocols. The diagram shows **(*top*)** the preparation (steps 1-4) and running (steps 5-7) of AMBER GB/SA implicit solvent simulations for a ~300 bp supercoiled DNA minicircle, and **(*bottom*)** the preparation (steps 8-15) and running (steps 16-18) steps of GROMACS explicit solvent simulation of a ~300 bp supercoiled DNA minicircle. The flow diagram also shows the types of analysis performed on both the implicitly and explicitly solvated MD trajectories. A conditional statement is made for the size of DNA minicircle. If the minicircle sequence is shorter than 150 basepairs, the implicit solvent GB/SA simulations need not be performed (skip steps 5-9) as the DNA cannot writhe.



**Figure 2** **a)** Schematic diagram showing the construction of an *N*-basepair planar circular DNA structure using the NAB program ‘circ.nab’: (left) the first basepair is placed at the circumference of a circle of radius *R* = (3.38 Å) / 2sin(π/*N*), and is oriented perpendicular to the tangent line, (middle) the system is rotated by an angle α = 2π / *N*. The second basepair is then added to replace the first basepair and is rotated by the twist angle θ = 2π×Lk / *N* and (right) the same process is repeated until a full DNA circle is made. **b)** Hybrid implicit/explicit solvation: (left-middle) A planar circular 336 bp DNA minicircle at ΔLk = -2 adopted a writhed conformation in AMBER GB/SA implicit solvent. (right) A selected writhed structure was then solvated in 0.1M NaCl and TIP3P water box by using GROMACS. **c)** Examples of DNA defects occurred when subjected to high bending and superhelical stress.



**Figure 3** **a)** Three sample DNA minicircle structures of the ‘opened’, ‘figure-8’ and ‘handcuff’ conformation. The ‘opened’ minicircle possesses the largest radius of gyration compared to the other two, while the ‘handcuff’ minicircle has the smallest radius of gyration and is the most compact structure. **b)** A writhed DNA minicircle structure viewed from two different angles, showing the change in the number of apparent crossing points from different viewpoints. **c)** DNA writhing calculation: a helical path (blue) is extracted from the atomistic structure of a minicircle DNA. The smoothed central helical axis (red) is then calculated by using ‘WrLINE.py’ python script. From this helical axis path *s* writhe can be calculated by integration of the coiling of *s* about itself (*s’*).



**Figure 4 a)** Time averages of writhe (purple) and radius of gyration (maroon) over the last 10 ns of the implicitly solvated DNA minicircles at seven different ΔLks. Orange marks indicate the relative mobility (down the vertical axis) of the seven minicircle topoisomers determined by gel electrophoresis experiments ***(19)***. **b)** An atomistic structure of a ΔLk = -2 DNA minicircle, obtained from an explicity solvated MD simulations. The structure is superimposed into a cryo-ET density map and visualised by VMD.

## Tables

|  |  |  |  |
| --- | --- | --- | --- |
| **Simulations** | **108 bp Explicit** | **336 bp Implicit** | **336 bp Explicit** |
| GB/SA  Number of atoms  Number of nucleotide residues  Number of water residues  Number of ionic residues  HPC system  Number of processors used  Speed (ns/day) | No  ~ 450,000  216  ~ 150,000  ~ 750  ARC1/ARC2  32  ~ 1 | Yes  21,373  672  -  -  ARC1/ARC2  32  ~ 1 | No  ~ 2,100,000  672  ~ 700,000  ~ 3,200  ARCHER  256  ~ 5 |

**Table 1** Comparison between implicit and explicit solvation of a 336 bp DNA minicircle system for the use of generalised born Surface approximation (GB/SA), numbers of atoms, numbers of nucleotide residues, numbers of water residues, numbers of ionic residues, numbers of processors working in parallel and simulation speed in ns/day.

|  |  |
| --- | --- |
| **Categories** | **Files** |
| a) AMBER GB/SA preparation | * ‘circ.nab’: create a PDB file of DNA minicircle * ‘leapscript\_imp’: generate AMBER topology and starting coordinate files * ‘leaprc.ff99circ’: a forcefield parameter set specially used for DNA minicircles (DNA with no ends) * ‘all\_nuc94bsc0\_chiOL4.in’, ‘frcmod.parmbsc0’ and ‘frcmod.OL4.chi’: forcefield modification files for the dihedral parameters * ‘dangions.dat’: a forcefield modification file for the ion electrostatic parameter |
| b) AMBER GB/SA simulation | * ‘molecule.prmtop’ and ‘molecule.inpcrd’: AMBER topology and starting coordinate files * ‘min1.in’ and ‘min2.in’: input parameters for minimisation in GB/SA implicit solvent * ‘md1.in’, ‘md2.in’ and ‘md3.in’: input parameters for equilibration in GB/SA implicit solvent with all atom coordinate restraints * ‘md4.in’: input parameters for a productive MD run in GB/SA implicit solvent with NMR distance restraints on hydrogen bonds * ‘RST’: all the information on the NMR restraints * ‘gbsa.sh’: shell commands to execute the minimisation, equilibration and productive MD runs within AMBER GB/SA implicit solvent |
| c) GROMACS explicit solvent preparation | * ‘amb2gmx\_dihe.pl’: perl script to convert the AMBER topology and coordinate files into GROMACS format * ‘ffbsc0.itp’, ‘tip3p.itp’ and ‘ions.itp’: ff99bsc0 forcefield for DNA, TIP3P water and ion models in GROMACS topology file format * ‘molecule.w.top’, ‘molecule.wp.top’, and ‘molecule.wnr.top’: GROMACS topology files for the system with water, counter-ions and monovalent salt ions under no artificial restraints |
| d) GROMACS explicit solvent simulation | * ‘molecule.wpr.top’: the GROMACS topology file for the system with water and ions under coordinate position restraints * ‘em.mdp’, ‘eq.mdp’ and ‘md.mdp’: input parameter files for minimisation, equilibration and productive MD runs * ‘posre.itp’: a topology file containing information of coordinate position restraints * ‘gmxmd.sh’: shell commands to execute the minimisation, equilibration and productive MD runs within GRAMACS explicit solvent * ‘xtc2crd.tcl’: a tcl script for the VMD program to convert a GROMACS \*.xtc compressed trajectory file into the AMBER \*.mdcrd format |
| e) WrLINE | * ‘WrLINE.py’: to run a package of PTRAJ and python scripts for the writhe measurement from AMBER topology and coordinate files |

**Table 2** List of the files used in a) preparing the minicircle DNA structures for GB/SA simulations, b) running the AMBER GB/SA implicit solvent simulations, c) preparing the GROMACS explicit solvent simulation, d) running the GROMACS explicit solvent simulations and e) measuring writhe values by WrLINE.

|  |  |  |
| --- | --- | --- |
| **Stages** | **Restraints** | **Durations** |
| **Mininisation 1** | k = 50.0 kCal/mol/Å2 | 10000 cycles |
| **Mininisation 2** | No restraints | 10000 cycles |
| **Equilibration 1** | k = 50.0 kCal/mol/Å2, T=100-300K | 10 ps |
| **Equilibration 2** | k = 10.0 kCal/mol/Å2, T = 300K | 100 ps |
| **Equilibration 3** | k = 1.0 kCal/mol/Å2, T = 300K | 200 ps |
| **Productive Runs** | ***NMR***, k = 1.0 kCal/mol/Å2 | 10-100 ns |

**Table 3** Force constants and time durations used in each stage of the AMBER GB/SA implicit solvent simulation protocols

|  |  |  |
| --- | --- | --- |
| **Stages** | **Restraints** | **Durations** |
| **Mininisation 1** | k = 500.0 kCal/mol/Å2 | 10000 cycles |
| **Mininisation 2** | k = 50.0 kCal/mol/Å2 | 10000 cycles |
| **Mininisation 3** | k = 25.0 kCal/mol/Å2 | 10000 cycles |
| **Mininisation 4** | No restraints | 10000 cycles |
| **Equilibration 1** | k = 500.0 kCal/mol/Å2, T = 100K | 10 ps |
| **Equilibration 2** | k = 50.0 kCal/mol/Å2, T=100-300K | 10 ps |
| **Equilibration 3** | k = 50.0 kCal/mol/Å2, T = 300K | 10 ps |
| **Equilibration 4** | k = 25.0 kCal/mol/Å2, T = 300K | 10 ps |
| **Equilibration 5** | k = 10.0 kCal/mol/Å2, T = 300K | 10 ps |
| **Equilibration 6** | k = 5.0 kCal/mol/Å2, T = 300K | 10 ps |
| **Equilibration 7** | k = 2.5 kCal/mol/Å2, T = 300K | 10 ps |
| **Equilibration 8**  **Productive Runs** | k = 1.0 kCal/mol/Å2, T = 300K  (none) | 10 ps  10-100 ns |

**Table 4** Force constants and time durations used in each stage of the GROMACS explicit solvent simulation protocols