ISSN 1674-1056
http://cpb.iphy.ac.cn
http://iopscience.iop.org/cpb
中国物理 B
Chinese Physics
中国物理
Chinese Physics B

Ethylene glycol solution-induced DNA conformational transitions

Nan Zhang(张楠), Ming-Ru Li(李明儒), Feng-Shou Zhang(张丰收)

Citation: Chin. Phys. B . 2018, 27(11): 113102. **doi:** 10.1088/1674-1056/27/113102

Journal homepage: http://cpb.iphy.ac.cn; http://iopscience.iop.org/cpb

What follows is a list of articles you may be interested in

A simulation study of water property changes using geometrical alteration in SPC/E

Ming-Ru Li(李明儒), Nan Zhang(张楠), Feng-Shou Zhang(张丰收)
Chin. Phys. B . 2018, 27(8): 083103. **doi:** 10.1088/1674-1056/27/8/083103

Diffusion behavior of hydrogen isotopes in tungsten revisited by molecular dynamics simulations

Mingjie Qiu(丘明杰), Lei Zhai(翟磊), Jiechao Cui(崔节超), Baoqin Fu(付宝勤), Min Li(李敏), Qing Hou(侯氢)
Chin. Phys. B . 2018, 27(7): 073103. **doi:** 10.1088/1674-1056/27/7/073103

Anisotropic self-diffusion of fluorinated poly(methacrylate) in metal-organic frameworks assessed with molecular dynamics simulation

Tao Lu(鲁桃), Biao Xu(徐彪), Fei-Hong Ye(叶飞宏), Xin-Hui Zhou(周馨慧), Yun-Qing Lu(陆云清), Jin Wang(王瑾)
Chin. Phys. B . 2017, 26(12): 123104. **doi:** 10.1088/1674-1056/26/12/123104

Effect of grain boundary structures on the behavior of He defects in Ni: An atomistic study

H F Gong(龚恒风), Y Yan(严岩), X S Zhang(张显生), W Lv(吕伟), T Liu(刘彤), Q S Ren(任啟森)
Chin. Phys. B . 2017, 26(9): 093104. **doi:** 10.1088/1674-1056/26/9/093104

Wetting and coalescence of the liquid metal on the metal substrate

Zhen-Yang Zhao(赵珍阳), Tao Li(李涛), Yun-Rui Duan(段云瑞), Zhi-Chao Wang(王志超), Hui Li(李辉)
Chin. Phys. B . 2017, 26(8): 083104. **doi:** 0.1088/1674-1056/26/8/083104

Ethylene glycol solution-induced DNA conformational transitions*

Nan Zhang(张楠)[1], Ming-Ru Li(李明儒)[2], and Feng-Shou Zhang(张丰收)[1,2,3,†]

[1] *Key Laboratory of Beam Technology of Ministry of Education, College of Nuclear Science and Technology, Beijing Normal University, Beijing* 100875, *China*
[2] *Beijing Radiation Center, Beijing* 100875, *China*
[3] *Center of Theoretical Nuclear Physics, National Laboratory of Heavy Ion Accelerator of Lanzhou, Lanzhou* 730000, *China*

(Received 10 July 2018; revised manuscript received 26 August 2018; published online 24 October 2018)

We study the properties of the ethylene glycol solutions and the conformational transitions of DNA segment in the ethylene glycol solutions by molecular dynamics simulations based on GROMACS. The hydrogen network structures of water–water and ethylene glycol–water are reinforced by ethylene glycol molecules when the concentrations of the solutions increase from 0% to 80%. As illustrated by the results, conformation of the double-stranded DNA in ethylene glycol solutions, although more compact, is similar to the structure of DNA in the aqueous solutions. In contrast, the DNA structure is an A–B hybrid state in the ethanol/water mixed solution. A fraying of terminal base-pairs is observed for the terminal cytosine. The ethylene glycol molecules preferentially form a ring structure around the phosphate groups to influence DNA conformation by hydrogen interactions, while water molecules tend to reside in the grooves. The repulsion between the negatively charged phosphate groups is screened by ethylene glycol molecules, preventing the repulsion from unwinding and extending the helix and thus making the conformation of DNA more compact.

Keywords: DNA conformational transitions, ethylene glycol, molecular dynamics simulations

PACS: 31.15.xv, 36.20.Ey, 36.20.Hb **DOI:** 10.1088/1674-1056/27/11/113102

1. Introduction

Water is the most important life supporting solvent.[1,2] Two thirds of the environment surrounding the DNA molecules in the cell nucleus is composed of water molecules.[3,4] This provides an appropriate environment for both biophysical and biochemical processes[5–7] to develop. A basic question of interest is which fundamental property of its molecule distinguishes water as the most important solvent in nature.[8,9] Are there any other liquids that support biological activity? An important property of water is that it dissolves charged ions and polar molecules easily. It is generally considered that polarity and hydrogen bonding are the distinguishing attributes of water when compared with other liquid.[10,11]

Many experimental and theoretical studies have reported[12–18] the properties of DNA macro-molecule in neutral aqueous. There are many kinds of solvents, such as acids or alkaline solvents, organic or inorganic solvents, and hydrophilic or hydrophobic solvents. It is ostensibly undesirable that strong acids and strong alkaline solutions are corrosive, or DNA is insoluble in most organic solvents. An exception is the organic solvent ethylene glycol. It has been observed that DNA can retain its native double-stranded structure in ethylene glycol–salt solutions.[19,20] The biological activity of DNA in ethylene glycol has been investigated by following the transforming activity of bacillus subtilis DNA with competent bacteria.[21] DNA can keep its biological activity after long time periods in ethylene glycol–salt solutions. However, DNA in ethylene glycol has been little investigated during last half century.

In addition, ethylene glycol solutions have the following advantages in comparison with aqueous solutions.

(i) It is known that radiation damage ionizes the water molecule and creates hydroxyl radicals[22–25] which adversely affect a cell's DNA duplexes. It seems possible that DNA might be more resistant to radiation damage in ethylene glycol than in water, since hydroxyl radicals would not be generated in similar amounts in irradiated ethylene glycol solutions. The randomly distributed clusters of diatomic OH-radicals that are the primary products of megavoltage ionizing radiation in water-based systems are the main source of hydrogen-abstraction as well as formation of carbonyl and hydroxyl groups in the sugar moiety that create holes in the sugar rings. These holes grow slowly between DNA bases and DNA backbone with the damage affecting both DNA single and double strand breaks.[26,27] Various effects of the ionizing radiation on biological systems range from the development of genetic aberrations, carcinogenesis to aging.[28,29] Ethylene glycol may provide better protection for DNA against cosmic radiation, hence facilitating its safe interplanetary transport.

(ii) It is possible to work with double-stranded DNA in ethylene glycol solutions in the convenient temperature range of -10 °C to $+25$ °C, where micro-organisms do not usually grow.[19] Thus the ethylene glycol may present a cheaper and more efficient option for storing DNA in bio-banks than in

*Project supported by the National Natural Science Foundation of China (Grant Nos. 11635003, 11025524, and 11161130520), the National Basic Research Program of China (Grant No. 2010CB832903), and the European Commissions 7th Framework Programme (FP7-PEOPLE-2010-IRSES) (Grant No. 269131).

†Corresponding author. E-mail: fszhang@bnu.edu.cn

© 2018 Chinese Physical Society and IOP Publishing Ltd

frozen aqueous buffer solutions.

Computational modeling can help to gain a detailed understanding of DNA structure and function. The most accurate approach would encompass the use of the quantum mechanical (QM) wave function. However, realistic modeling of DNA molecule and its environment requires extensive computing resources and can only be applied to a limited number of conformations. Therefore, most simulation projects employ classical molecular mechanics (MM) energy functions. Because of the need for high statistical computation of these functions during the simulation, the force fields used are relatively simple and utilize many adjustable empirical parameters which are generated by fitting the experimental properties or quantum chemical calculations.

Theoretically, the properties of ethylene glycol solutions and the conformational transitions of DNA are all determined by the microscopic interactions between the solvents and DNA molecules. In various experimental methods (x-ray crystallography, neutron diffraction, nuclear magnetic resonance (NMR) and fluorescent techniques, and atomic force microscope) and theoretical methods, the method of the molecular dynamics (MD) simulation can help us understand many macroscopic and microscopic phenomena of inter- and intra-molecular interactions, and simultaneously pave a useful way of discovering flexibility and tracing conformation transitions at atomic level[30–35] which would complement the static structure imaging techniques and help interpret experimental results in structure studies. Using MD simulation method with different parameters, we can study the effect of ethylene glycol on the hydrodynamic properties of DNA in solution. Their results are helpful for understanding the properties of DNA in ethylene glycol solutions with applications in radiation protection and DNA store.

In this paper, a series of water and ethylene glycol mixture solvents with five different weight concentrations of ethylene glycol are designed by the PACKMOL code package.[36] Conformational transitions of the duplex d(CGCGAATTCGCG) in ethylene glycol are studied at 293 K temperature using MD simulation. As concentration rises, the hydrogen network between water–water and ethylene glycol–water changes. Ethylene glycol then forms hydrogen bond with the negatively charged oxygen atoms of phosphate groups around the backbone.

2. Model and methods

2.1. DNA and solvent models

The Drew–Dickerson dodecamer (DDD) has been chosen for the initial state of DNA (PDB ID: 171d). The reason for this choice is that (i) DDD contains the recognition site of the EcoRI restriction enzyme and is frequently used in gene recombination techniques,[37] and (ii) it has been extensively studied experimentally and in nanosecond-to-microsecond MD simulations.[38–40] In addition, the DNA double strand is long enough to maintain a complete B-helix so that we can make the model more representative of a longer double-stranded DNA sequence. Stereo views of the initial 171d are given in Fig. 1.

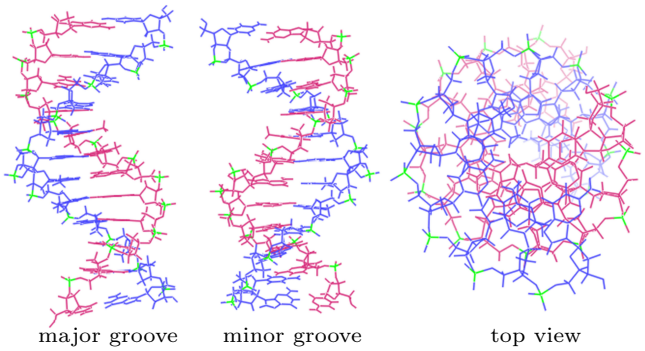


Fig. 1. (color online) Stereo views of the initial 171d structure. Presented here are 3 typical views from major groove, minor groove, and top. The systems are treated with the PARMBCS1 force field.

Although some sophisticated potential functions of water are still being developed,[8] in this work the most widely used TIP3P model is adopted because its description is, generally, satisfactory under ambient conditions when compared with experiments.[41] PARMBSC1 force fields are used to describe oligonucleotide interactions in ethylene glycol solutions (Table 1 and Fig. 2).[42] Electrostatic interactions are dealt with by the particle mesh Ewald summation method using a cutoff of 1.2 nm for the interaction range, with the real-space cutoff set at 1.0 nm[43,44] and the Lennard–Jones interactions cutoff set at 1.0 nm. LINCS is used to restrain all chemical bonds involving hydrogen atom.[45]

Table 1. The force field parameter for ethylene glycol models in molecular dynamics simulations.

Atom	σ/Å	ε/kJ·mol^{-1}	Mass/a.u.	Charge/$\lvert e\rvert$
C1	3.3997	0.4577	12.0100	0.298635
C2	3.3997	0.4577	12.0100	0.298635
H1	2.4714	0.6569	1.0080	0.000348
H2	2.4714	0.6569	1.0080	0.000348
H3	2.4714	0.6569	1.0080	0.000348
H4	2.4714	0.6569	1.0080	0.000348
O1	3.0665	0.8803	16.0000	−0.737693
O2	3.0665	0.8803	16.0000	−0.737693
H5	0.0000	0.0000	1.0080	0.438361
H6	0.0000	0.0000	1.0080	0.438361

The intermolecular energy consists of two parts: $E = E_{\mathrm{LJ}} + E_{\mathrm{C}}$, where E_{LJ} is the Lennard–Jones interaction and E_{C} is the electrostatic interaction. They are expressed as

$$E_{\mathrm{LJ}} = 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right], \tag{1}$$

$$E_{\mathrm{C}} = \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}, \tag{2}$$

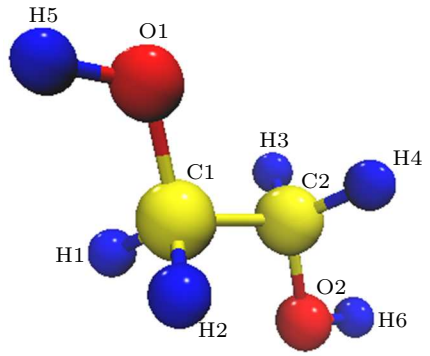


Fig. 2. (color online) Stereo views of ethylene glycol. The red, yellow, and blue represent oxygen, carbon, and hydrogen atom, respectively.

where ε_0 is the vacuum permittivity and r_{ij} presents the distance between two atoms. The Lennard–Jones core provides short-range inter-molecule repulsion and describes the molecular size.[46] The Coulomb interactions between point charges describe the effects of hydrogen bond and give the local orientational structure.[47]

2.2. Simulation description

Classical MD simulations have been performed based on GROMACS code packages to study the conformational transitions of DNA-molecule in the ethylene glycol solution.[48] One DNA dodecamer, 22 Na^+ counter-ions, and 8040 solvent molecules (water and ethylene glycol, big enough to ensure that the DNA does not interact with its periodic images) are contained in a periodic cubic box. The aim of this work is to study the solvent effects on DNA conformation. Two groups of parallel simulations are conducted.

The first one is the simulation of the behavior of aqueous ethylene glycol mixtures at 293 K. The data cover the entire range of composition at 20% mass intervals. The time step is 2 fs and the total simulation time is 200 ns. A complete overview of the simulation systems is given in Table 2.

After that, 171d segment has been put into the concentration of solvents 40%, with 22 Na^+ counterions for each system. The time step is 2 fs and the total simulation time is 200 ns.

Each of these molecular dynamics simulations consists of four stages. At the beginning, we employ the energy minimization to ensure that the system has no steric clashes or inappropriate geometry. Then the system is thermalized and pre-equilibrated using our standard multi-step protocol[50] followed by 1 ns of equilibration. Next, an isothermal–isobaric ensemble ($P = 1$ atm and $T = 293$ K) uses the Berendsen algorithm[51] to control the temperature and the pressure, with a coupling constant of 5 ps. Finally, an MD production run of 200 ns is performed to provide configurations for analysis.

Table 2. Overview of ethylene glycol/water simulations: number of molecules, ethylene glycol mass fraction C_i, mole fraction X_i, and fractional error $\delta(\rho)$. Experimental densities from Ref. [49] are the equilibrium properties of the mixtures.

H_2O	$(CH_2OH)_2$	C_i	X_i	$\rho_{exp.}$(293.15 K)	$\rho_{theo.}$(293 K)	$\delta(\rho)/\%$
8040	0	0%	0.0000	998.22	991.22	0.70
7500	540	20%	0.0677	1024.52	1025.19	0.00
6732	1308	40%	0.1622	1051.78	1061.92	0.96
5604	2436	60%	0.3003	1076.57	1099.56	2.14
3720	4320	80%	0.5373	1097.84	1138.45	3.70

2.3. DNA structure parameters

Structure parameters describing the axis-base pair (X-displacement, Y-displacement, inclination), intra-base pair (shear, stretch, stagger, buckle, propeller, opening), inter-base pair (rise, roll, shift, slide, tilt, twist), backbone torsion (BI/BII population, canonical α–γ, puckering), and groove characters (MW, mW, MD, mD) of DNA are analyzed by programs Curves+ and Canal.[52] The X-displacement (Xdisp) describes the displacement of a base pair along the X direction (the direction of the short axis of the base pair) of the base pair axis. Inclination (Incl) is defined as dihedral angle between the base pairs and the plane with the initial x and y axis. Twist and roll ρ are relative rotation around the z axis and rotation around the y axis between two successive base pairs in an overall helix axis, respectively.[53] The width (MW for major groove and mW for minor groove) and depth (MD for major groove and mD for minor groove) of the groove are defined in Ref. [53]. All of these parameters are analyzed continually for the precise values in our simulations. The B-DNA conformation can most easily be distinguished from the A conformation with these parameters.

The data for the DNA in pure water and ethanol solutions are from the BIGNASim database (http://mmb.irbbarcelona.org/BigNASim/index.php).

3. Results and discussion

3.1. Properties of ethylene glycol solutions

Firstly, we demonstrate the information correctly on the molecular interactions between water and ethylene glycol in binary mixtures at a series of weight percent of ethylene glycol. The calculated results are compared with the experimental data. For easy comparison, the simulated densities are tabulated in Table 2 alongside experimental densities[49] at 293 K. The theoretical value of the density in pure aqueous solution is less than the experimental value. However, after adding the ethylene glycol solutions, it increases to a higher value than the experimental one.

In the mixtures solvent of 20% (Table 2), the average error for the densities is 0.67 kg/m^3 (relative error near 0.00%), whereas it is 40.61 kg/m^3 (relative error 3.70%) with 80% percent. The reason for the density results may be that the interactions between water and ethylene glycol molecules disturb the original water structure, while the TIP3P model used in the simulation is rigid implying that the structure change of water molecular has not been reflected. When the weight percent of ethylene glycol rises from 0% to 80%, the average error for the densities is also increased. In different concentrations of ethylene glycol solutions, the maximum value of the error ($C_i = 80\%$) is less than 4%. Taking into account the required extensive computer resources and comparing with experimental results, the TIP3P water model is good enough to study the conformation transition of DNA.

Figure 3 shows the radial distribution functions (RDFs) of water (oxygen (ow) and hydrogen (hw)) and ethylene glycol (oxygen (o1)) molecules in a series of concentrations of ethylene glycol solutions at room temperature 293 K. The water–water RDFs are presented in Figs. 3(a)–3(c) and a comparison of water–water correlations with those of pure water shows that the first peak value increases as the ethylene glycol concentration increases. Comparison with pure water shows that whether the mixture is water-rich or not, the interactions between water molecules are not destroyed by the ethylene glycol molecules as there are still hydrogen bonding interactions between water molecules. As the concentrations increase, the first peak enhances, suggesting that there is a tendency for the association of a small number of water molecules in this system.

Results for ethylene glycol–water RDFs are shown in Figs. 3(d)–3(e). The first peaks of $g(r)_{\mathrm{o1-ow}}$ and $g(r)_{\mathrm{o1-hw}}$ occur at 0.28 nm and 0.18 nm, respectively, revealing the presence of hydrogen bonding in ethylene glycol–water molecules. As can be seen, the first peak values of these RDFs are also increased as the weight percent of ethylene glycol increases from 0% to 80%, indicating that the hydrogen-bonded network structure strengthens between ethylene glycol–water. Comparison of $g(r)_{\mathrm{ow-hw}}$ Fig. 3(b) and $g(r)_{\mathrm{o1-hw}}$ Fig. 3(e) shows that the hydrogen bond acceptability of the oxygen in the ethylene glycol is less than the oxygen in the water.

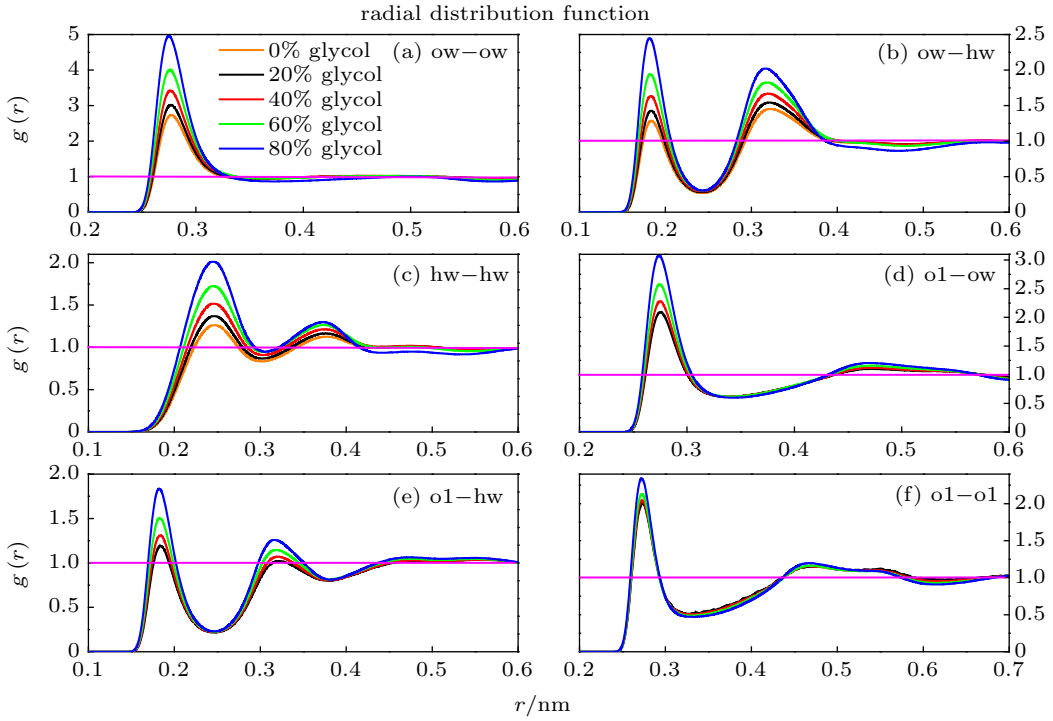


Fig. 3. (color online) The radial distribution functions of water and ethylene glycol in a series of concentrations glycol solutions. (a) water oxygen (ow) –ow, (b) ow–water hydrogen (hw), (c) hw–hw, (d) ethylene glycol oxygen (o1) –ow, (e) o1–hw, and (f) o1–o1. Mass fractions 0%, 20%, 40%, 60%, and 80% are shown as orange, black, red, green, and blue, respectively.

The behavior exhibited by the first peaks may be explained by the spatial density functions (SDFs) of water around one ethylene glycol molecule as indicated in Fig. 4. Two rings can be seen around the acceptor cap in front of the hydroxyl oxygen, which are due to the oxygen atom in the ethylene glycol acting as an acceptor and forming hydrogen bond with the hydrogen atoms in the water. The other distinct feature is the bowl appearing below the ethylene glycol molecule, and the hydrogen atom in the ethylene glycol can be used as a donor attracted by oxygen atoms in water since an ethylene glycol molecule can, on average, form 4 hydrogen bonds (Fig. 2). The hydrogen-bonded network structures between water and water and between ethylene glycol and water are strengthened as the weight percent of ethylene glycol in-

creases.

The first peak of $g(r)_{\mathrm{o1-o1}}$ occurs at 0.28 nm, similar to that exhibited in water–water $g(r)_{\mathrm{ow-ow}}$, revealing that the ethylene glycol–ethylene glycol hydrogen bonding persists, even as this component becomes more diluted. An examination of Figs. 3(a), 3(d), and 3(f) clearly shows that water prefers to hydrogen bond to itself. However, ethylene glycol prefers to form hydrogen bond with water rather than with itself. Note that the distribution of the first shell molecules between the ethylene glycol–ethylene glycol molecules is not obviously affected by the concentration of the ethylene glycol solutions. Since ethylene glycol prefers to form hydrogen bond with water rather than with itself, the water molecules form a hydrated shell around the ethylene glycol molecule to screen the ethylene glycol molecules. Furthermore, addition of ethylene glycol amplifies the H bond network structure of ethylene glycol–water, while the distribution between ethylene glycol molecules is obviously not affected by the concentration of the ethylene glycol solutions.

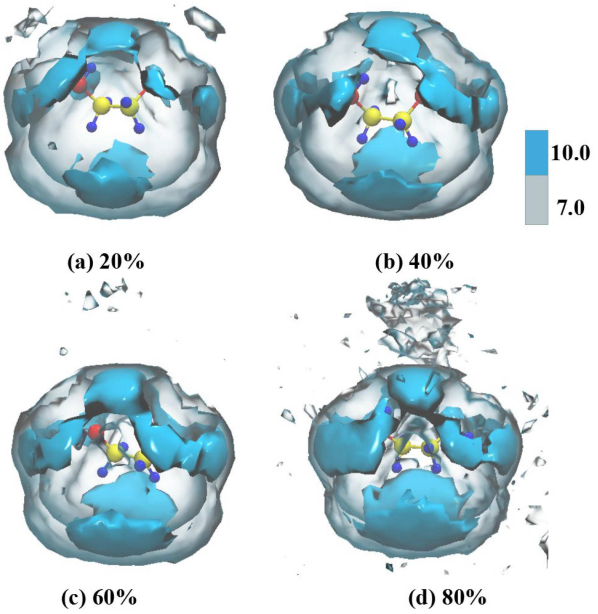


Fig. 4. (color online) The spatial density functions of water around one ethylene glycol molecule in (a) 20%, (b) 40%, (c) 60%, and (d) 80% ethylene glycol solvents. The scale of the isosurface is the relative density of water. The red, yellow, and blue represent oxygen, carbon, and hydrogen atom, respectively.

The changes of hydrogen-bonded network structure in these solutions can be presented in Fig. 5 vividly with spatial density functions of water around one water molecule, which gives the rendering of relative positions and probability of neighboring water molecules around the central molecular. Here the average density of water is set as 1.0, with the isosurfaces of relative density of water $\rho_{\mathrm{r}} = 15$, 10, and 7 presented. Figure 5(a) shows that the SDFs of pure water solution in combination with isosurfaces of $\rho_{\mathrm{r}} = 15$ and 10 give the most favorable tetrahedral coordinate positions around one H_2O molecule. The two prominent caps directly interact with the H atoms of the central water molecule, which are due to its hydrogen-bond-accepting by neighboring water molecules. The single large feature below the central oxygen molecule is due to its two hydrogen-bond-donating by neighboring water molecules. The distribution of $\rho_{\mathrm{r}} = 7$ is mainly determined by the shape of the second hydration shell of water. In pure water, the profile of $\rho_{\mathrm{r}} = 7$ is like two thin rings around each hydrogen atom and perpendicular to the O–H bond respectively. Also two shallow caps outside of the more prominent caps are observed.

Comparing effects of different concentrations shows that the $\rho_{\mathrm{r}} = 15$ and 10 isosurfaces occupy large areas in high concentrations solutions. This is particularly noticeable for the nearest neighbors compared to the central oxygen site. As the concentration of the solutions increases, the second hydrogen shell of water is destroyed. This effect is seen in Fig. 5(e) where the thin rings and shallow caps have disappeared while the prominent caps and the large feature below the central oxygen molecule become thicker and broader. This is due to that the water–water structure becomes denser, shown in RDFs, with increasing concentrations of ethylene glycol solutions.

The information on the molecular interactions between water and ethylene glycol at a series of concentrations is important for our understanding of complex interactions between DNA and ethylene glycol solutions. Here we present the change of H-bonded network structure induced by concentration change through some distributions functions such as radial distribution functions and spacial density functions. Firstly, revealing the presence of hydrogen bonding in

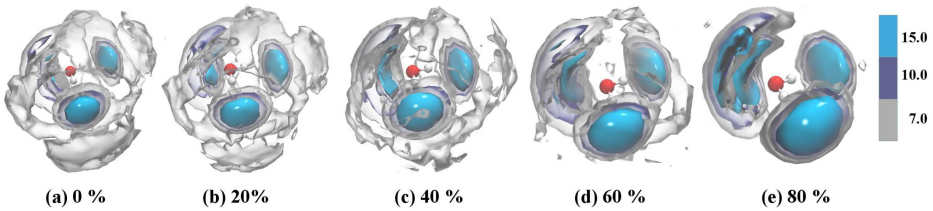


Fig. 5. (color online) The spatial density functions of water around one water molecule in (a) 0%, (b) 20%, (c) 40%, (d) 60%, and (e) 80% ethylene glycol solvents. The scale of the isosurface is the relative density of water. The red and blue represent oxygen and hydrogen atom, respectively.

water–water, ethylene glycol–water, and ethylene glycol–ethylene glycol, the intermolecular hydrogen bonding combinations have the following priority: water–water > ethylene glycol–water > ethylene glycol–ethylene glycol. As the hydrogen bonded network structure is affected by the concentration, the first shell layer between water–water and ethylene glycol–water molecules increase. On the other hand, the distribution of the first shell molecules between the ethylene glycol–ethylene glycol molecules is not obviously affected by the concentrations of the ethylene glycol solutions.

3.2. Conformational transitions of DNA

It was earlier shown that the transforming activity of DNA from bacillus subtilis was unaffected in salt-containing ethylene glycol solutions at room temperature. We now discuss the analysis of the 171d in ethylene glycol solutions. Molecular dynamics simulations were first carried out at 293 K to access the thermodynamic stability of the DNA structures in ethylene glycol solutions.

To study the thermodynamic stability of DNA conformation, the duplex DNA is put into the TIP3P model solvent with ethylene glycol ($C_i = 40$) at 293 K. At $C_i = 20$, it seems that the main structure features are similar to those observed for pure water. However, as the weight percent of ethylene glycol increases, the average error for the densities also increased. Thus the 40% weight percent of ethylene glycol is chosen as the solvent environment for the simulations of DNA.

The main results are presented in Fig. 6. Based on their root mean square deviations (RMSD), the DNA trajectories, overall, are stable and no major structural distortions are found. The average RMSDs of DNA trajectories with respect to initial structures, x-ray structures (PDB ID: 1BNA), and NMR structures (PDB ID: 1NAJ) are 0.28 (standard deviation 0.0369), 0.38, and 0.31 nm, respectively. Comparison with water/ethanol mixtures solvent indicates that in water solvent the RMSD with respect to the initial structures is 0.16 nm with 0.0323 standard deviation throughout the whole simulation. Similarly, in water/ethanol mixture solvent, the RMSD with respect to the initial structures is 0.35 nm with 0.0649 standard deviation. This means that in water solvent the structural differences between the crystal and MD simulation are the most smallest among the three solvents. In contrast, the fluctuations of RMSD are larger for water/ethanol mixture solvent than the other two.

In general, we observe that DNA can retain a stable structure in ethylene glycol solutions. A more detailed discussion of DNA conformational transitions is carried out below. We begin with a discussion of the impact on DNA structures by water/ethylene glycol solvents. We then turn to a discussion of the average structural parameters and a comparison with other solutions.

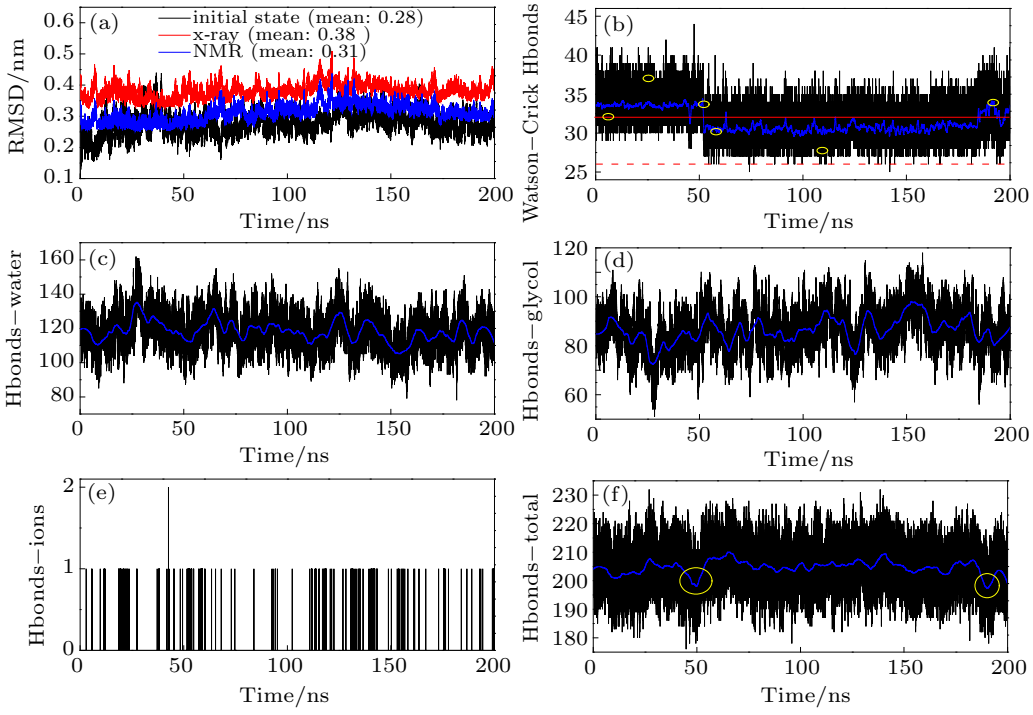


Fig. 6. (color online) (a) RMSD of the DNA trajectories with respect to the initial structure (black line), x-ray (red line), and NMR (blue line). (b) Evolution of the total number of Watson–Crick hydrogen bonds with time. Red solid is the ideal hydrogen bond number between the 12 base pairs while the red dashed line is the ideal hydrogen bond number without the terminal base pair. (c) Evolution of the total number of hydrogen bonds between water and DNA with time. (d) Evolution of the total number of hydrogen bonds between ethylene glycol and DNA with time. (e) Evolution of the total number of hydrogen bonds between ion and DNA with time. (f) Evolution of the total number of hydrogen bonds between solvents (water+ethylene glycol+ions) and DNA with time. The running averages over 1 ns are shown in blue line. The weight percent of ethylene glycol solution is 40%.

3.2.1. Solvent impacts on DNA

We first look at the evolution of the total number of Watson–Crick hydrogen bonds with time, shown in Fig. 6(b) by GROMACS. The red line represents the ideal state where all the 32 hydrogen bonds have formed in the 12 base pairs. The red dashed line represents a scenario in which 26 hydrogen bonds have formed in the inner 10 base pairs without considering the terminal base pairs.

In this paper, the hydrogen bonds are analyzed through two different programs, GROMACS and VMD. It is worth noting that the hydrogen bond determination criteria in different programs may not be consistent. In fact, there are multiple criteria for the determination of hydrogen bonds, such as energy criteria, electronic structure criteria, geometric criteria, and so on. The first two criteria are computationally intensive and can not be achieved with classical molecular dynamics simulations. Therefore, the geometric criteria is used in this manuscript. Even geometric criteria still have different forms. In this paper, the hydrogen bond is determined by the cut-off of the angle formed by the hydrogen donor–acceptor and the cut-off of the donor–acceptor distance. The standard for determining the presence of H-bonds is the distance between the acceptor and donor atom < 3.5 nm (corresponding to the first minimum position of the SPC water model RDF) and angle between the acceptor and donor atom $< 30°$.[54] The blue indicates the averages at 1 ns intervals.

Also shown are: (i) in the range 0–50 ns, the number of hydrogen bonds exceeding the ideal number, (ii) the number of hydrogen bonds being in the range 32–26 majority of time, and (iii) in the last 15 ns, the number of hydrogen bonds being higher than the ideal value. The number of hydrogen bonds in aqueous solutions are rarely found to be higher than the ideal value. Double-stranded DNA reversibly attains[55] a more compact and less extended conformation with a smaller solvent shell in ethylene glycol than in water. This structural feature may provide the possibility of more hydrogen bonds forming in the Watson–Crick strand.

To illustrate the hydrogen bond characteristics of DNA in the glycol solution, we present several conformations in Fig. 7 and Fig. 8 by VMD, corresponding to the number of characteristic hydrogen bonds in Fig. 6(b). As shown in Fig. 7(a), for $t = 5$ ns, the number of hydrogen bonds is 32. The main structural feature of DNA is the breakage of chains between 3C and 4G. There is no concept of bond breakage in classical molecular dynamics simulations, where we artificially consider phosphate bond breaks based on geometric criteria. If we use different hydrogen bond discrimination criteria, we may obtain a different number of hydrogen bonds. This conformation is short-lived and undergoes reversible transitions to the canonical structure. The details of the DNA chain breaks are shown in Fig. 7(b), where the phosphoric acid bond is broke between 3C and 4G. Figures 6(c) and 6(d) show the effect of phosphate bond breaks on neighboring base pairs, while from Fig. 6(c) it can be seen that the base pair 3C–22G may form six hydrogen bonds (the number of hydrogen bonds is 3 under normal circumstances). It is emphasized here that 6 hydrogen bonds are not formed between 3C–22G base pairs, but the distance between 3C–22G base pairs decreases so that the distances and angles between other hydrogen bond acceptor–donor sites satisfy the hydrogen bonding existence of geometric standards. This indicates that the distances between 3C–22G are more compact. In contrast, two hydrogen atoms are formed between 4G–21C, fewer than the normal number of three hydrogen bonds, pointing to larger 4G–21C distance (or angle).

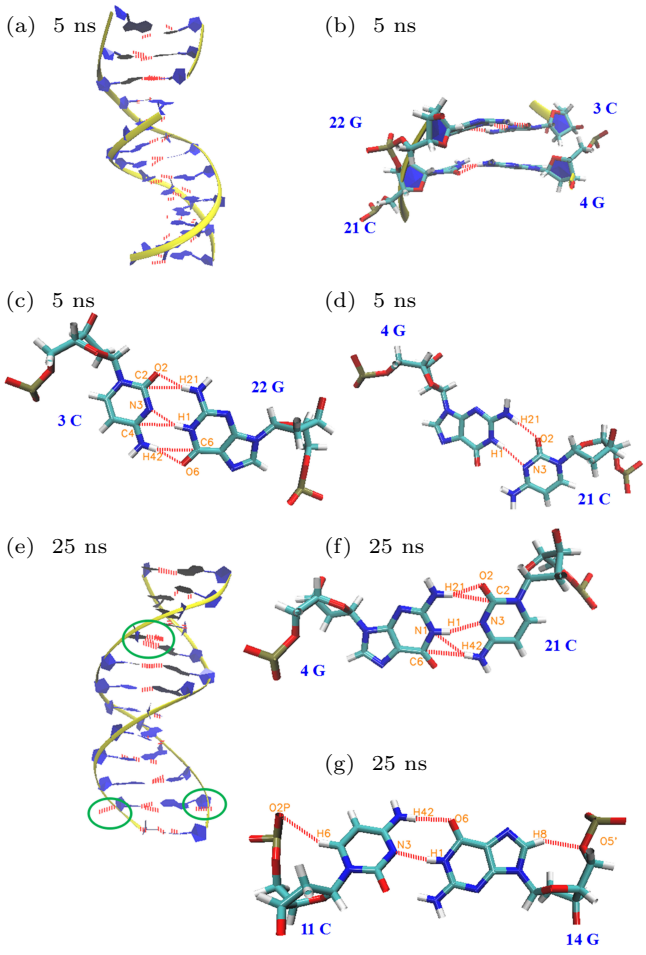


Fig. 7. (color online) Representative structures of the characteristics hydrogen bonds conformations found.

Figures 7(e)–7(f) show the two structural features of DNA at $t = 25$ ns. The first feature is the formation of five hydrogen

bonds between the base pairs of 4G–21C. The acceptor and donor atoms of hydrogen bond are marked in Fig. 7(f). Here we consider that not all five hydrogen bonds may form, but that the distance between donor and acceptor atoms provides the possibility of the presence of hydrogen bonds. It may be a statistical event. The second feature is characterized by a hydrogen bond break in the base pair of the 11C–14G, and a hydrogen atom on the sugar rings of C and G forms a hydrogen bond with the oxygen atom on the phosphoric acid to form a sub stable structure.

These structural features are short-lived and have no significant effect on the overall structure of DNA. A long-lived conformation is shown in Fig. 8.

The stepped shift, observed in the vicinity of 50 ns in Fig. 6, in the number of hydrogen bonds in the DNA is shown in Fig. 8. In Fig. 8(a), at $t = 51$ ns, before the stepped shift, there is no significant difference observed between the DNA conformation and the normal conformation. When $t = 54$ ns, after the stepped shift, the terminal cytosine (1C) undergoes fraying and flips out of the helix. At $t = 109$ ns, the cytosine that was turned out of the helix did not recover, and at the same times the phosphoric acid bond between 3C–4G is broken. At $t = 191$ ns, corresponding to the last hydrogen bond number increase in Fig. 6, the oxygen atoms on the sugar ring of the terminal cytosine have formed hydrogen bond with the hydrogen atoms on the sugar ring of 3G. In addition, the hydrogen atoms on the sugar ring of the terminal cytosine have formed hydrogen bond with the oxygen atoms on the sugar ring of 4C.

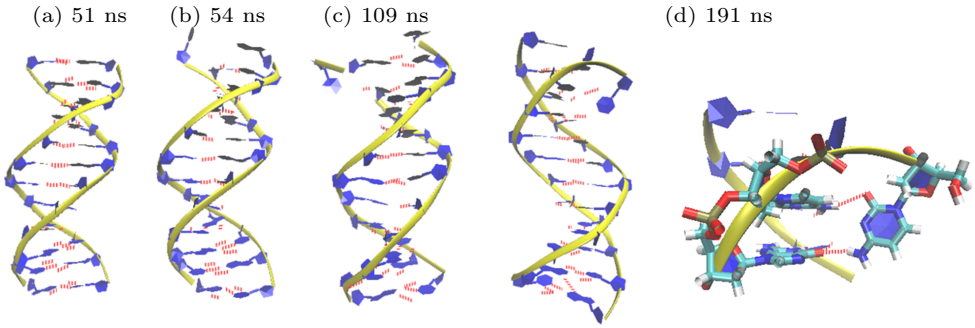


Fig. 8. (color online) Representative structures of the characteristic hydrogen bonds conformations found.

This dynamical feature is consistent with DNA in aqueous solution in previous MD simulations[38–40,56] and experiments.[57,58] However, in the case of aqueous solution, not only neither recent molecular dynamics simulations nor available experimental data support the presence of long-lived open state for terminal base pairs, but they reject the possibility of long-lived non-canonical conformations of terminal base pairs stabilized by interactions with the DNA grooves. We will discuss the reasons for this difference below.

Overall, there is a case for the number of hydrogen bonds exceeding the number of ideal hydrogen bonds for a long period which may be related to the more compact structure of the DNA in the ethylene glycol solution. A more compact structure may provide conducive conditions for the formation of hydrogen bonds in base pairs. We have found some short-lived non-canonical conformation, such as the phosphate acid bond break between the 3C and 4G and the sub state structure of 11C–14G base pair. We also found a long-lived non-canonical conformation, the fraying of the terminal base pair. Compared to their cytosine partner, no torsional flips are observed for the terminal guanine. The population of the non-canonical conformation is sequence-dependent, with lower probability on the central AATT.

We now look at the interaction of DNA molecules with solvents and counter-ions in solution. First of all, from Figs. 6(c) and 6(d), we can see that both water molecules and ethylene glycol molecules have hydrogen-bonding interactions with DNA. Second, the total number of hydrogen bonds formed by the solvent molecules with DNA is more oscillatory than the internal hydrogen bonds number of the DNA, whether water molecules or ethylene glycol molecules. That is to say, the hydrogen bonding between base pairs in DNA is more stable. The numbers of H-bonds–water and H-bonds–ethylene glycol indicate negative correlation, as shown in Figs. 6(c) and 6(d). With the increase of H-bonds–water, the number of H-bonds–ethylene glycol decreases, and vice versa. The calculated person correlation coefficient of 0.76 between H-bonds–water and H-bonds–ethylene indicates strong correlation. These phenomena pointed to competition between water and ethylene glycol molecules through the same interacting mechanism (hydrogen bonding).

Figure 6(e) shows the number of hydrogen bonds between DNA and sodium ions. Except in one case where hydrogen bonds is 2, during the 200 nanosecond simulation, all show 1 or 0 hydrogen bond forming. In fact, DNA tends to form electrostatic interactions with ions instead of hydrogen bond. H-bonds–ions and H-bonds–water (H-bonds–ethylene glycol) are not correlated to prove this point from the side. Finally,

figure 6(f) gives the total number of hydrogen bonds formed between the solvent (water + ethylene glycol + ions) and DNA where the two small peaks coincide with the stepped shift of the hydrogen bonds numbers (Fig. 6(b)). The fraying of the terminal base pairs also reduces the number of hydrogen bonds between the solvents and the DNA. However, the number of hydrogen bonds between the solvents and DNA molecule recovers and does not remain fixed as the cytosine flips.

The discussion of hydrogen bonds is summarized as follows. (i) There is a case where the number of hydrogen bonds between base pairs in DNA exceeds the number of ideal hydrogen bonds numbers. (ii) A long-lived terminal damage occurs. (iii) Both ethylene glycol molecules and water molecules interact with DNA through hydrogen bonds and there is a competition between them. (iv) Ions interact with DNA through strong chemical bonds.

It has been suggested that the physical and chemical mechanism of these DNA helices are relevant to solvent accessibility, base stacking interactions, phosphate hydration, and minor groove spine of hydration.[59–61] The most commonly concerned factor should be how the repulsion between negatively charged phosphate groups is screened.[62] Because the repulsion will unwind and extend the helix without screen and the different screen mechanism also affects the pipeline of the electrostatic repulsion, to check the underlying cause of the structural transitions induced by the solvent, we evaluate the water and ethylene glycol distribution around the phosphate groups on DNA helix.

The SDFs of the ethylene glycol molecules around the DNA are presented in Fig. 9(a). On the one hand, ethylene glycol molecules are distributed around the phosphate backbone to form rings structure that screens the repulsion between negatively charged phosphate groups. There is an indirect interaction between the phosphate oxygen atoms and ethylene glycol, acting as acceptor of phosphate oxygen atoms which form hydrogen bond with the hydrogen atoms in ethylene glycol. On the other hand, hydration ridge also forms in minor groove central base pairs AATT and almost no ethylene glycol molecular distribution is in major groove. A cap-shaped ethylene glycol distribution is also found at the DNA terminal. Figure 9(b) shows that water molecules in the major and minor grooves are distributed to support the backbone of DNA by forming hydration ridge. The sodium ions are distributed in the major and the minor grooves, and the difference in the distribution of water molecules is that the distribution of sodium ions in the major groove is regular. Sodium ions are only distributed around the nitrogen atoms and the oxygen atoms in the bases.

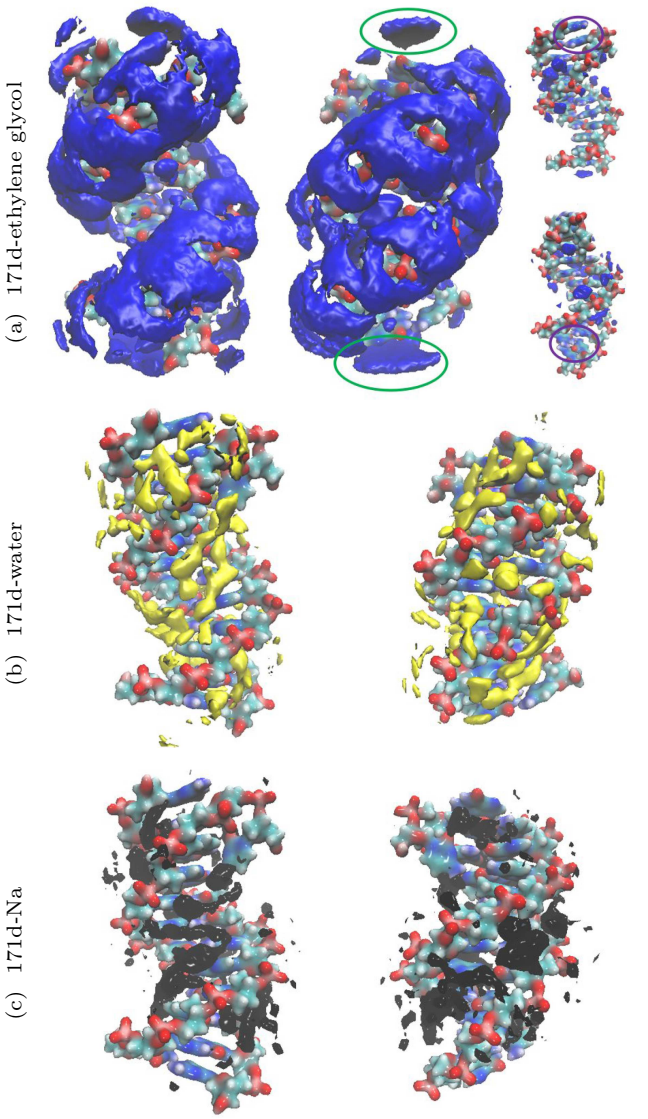


Fig. 9. (color online) (a) The spatial density functions of ethylene glycol molecules around DNA in ethylene glycol solvents. (b) The spatial density functions of water molecules around DNA in ethylene glycol solvents. (c) The spatial density functions of Na^+ around DNA in ethylene glycol solvents.

An interesting phenomenon observed, in contrast to the DNA in aqueous solution, is a ring structure of ethylene glycol molecules around the phosphoric acid backbone. This is, as previously stated above, due to the number of hydrogen bonds being higher than the ideal number in ethylene glycol solutions. Analysis of the spatial distribution functions shows that the repulsion between negatively charged phosphate groups is screened by ethylene glycol molecules and extends the helix. Herskovits *et al.* observed[55] that double-stranded DNA reversibly attains a more compact and less extended conformation with a smaller solvent shell in ethylene glycol than in water, while retaining intact biological activity. Because of the structure of DNA compact, it provides more possibility for hydrogen bonds in Watson–Crick strand. It is possible that the distribution of terminal cap-shaped ethylene glycol molecules causes long-life terminal fray.

3.2.2. Structural parameters

The structural parameters analysis is divided into the subsections: (i) backbone torsion, (ii) axis base pair, (iii) intra-base pair helical parameters, (iv) inter-base pair helical parameters, and (v) grooves.

The three major elements of flexibility in the backbone are (a) rotations around ζ/ε torsion, (b) rotations around α/γ torsion, and (c) sugar puckering.

The concerted rotation around ζ/ε torsion generates two major conformers: BI and BII , which are experimentally known to co-exist in an approximate ratio of 80%:20% (BI:BII) in B-DNA.[63] Figures 10(a)–10(c) show the BI /BII populations of base pairs in the Watson–Crick chain for the three solutions of ethylene glycol, ethanol, and water obtained by MD simulation giving average BI populations of 76%, 90%, and 78%, respectively. The populations of BI are significantly higher in ethanol than the other two solutions, while the populations of BI in water are higher than 2% of the ethylene glycol solution.

A comparison of the characteristics in Figs. 10(a)–10(c) with those of the aqueous solutions, as a standard, indicates that BII is mainly distributed on C or G bases, whereas the middle sequence AATT has very few BII distributions. In other words, the BI /BII probability distribution has a sequence dependency. In addition, a BII population peak appears at 4G (21C) step. Schwieters and Clore[64] performed an NMR experiment on Dickerson–Drew dodecamer together with large-angle solution x-ray scattering measurements and estimated the BII populations. The estimated BII populations of Schwieters and Clore and our MD data are at least qualitatively comparable, both showing a peak in the G4 step.

Analysis shows that in ethylene glycol solution, the BII population is higher than 90% in 4G step, which is obviously higher than the other two solutions. The location of this peak corresponds to the breaking point of the chain between 3C–4G, as discussed earlier. The populations of BII in ethanol solution are significantly lower than those in other two solutions, and higher than 30% in the middle base 7T.

On the basis of a statistical analysis of x-ray data, Djuranovic and Hartmann discovered that passing from BI/BI to mixed BII/BII substates implies a decrease in roll and increase in twist.[65] The data reported by Heddi *et al.*[66] indicate that slide increases, too. This is expected to be particularly important in understanding indirect recognition of DNA by proteins,[67,68] and also the mechanism of DNA intercalation by small organic compounds.[69]

Another structural deviation common in simulations is the flip of the backbone torsion angles α/γ from the canonical $g-/g+$ state to the non-canonical state as shown in Figs. 10(d)–10(f). The population of non-canonical α/γ torsion is not negligible. During the molecular dynamics simulation, the non-canonical α/γ torsion spends 4% of the time in the ethylene glycol solution, 10% in ethanol solution, and 3% in aqueous solution.

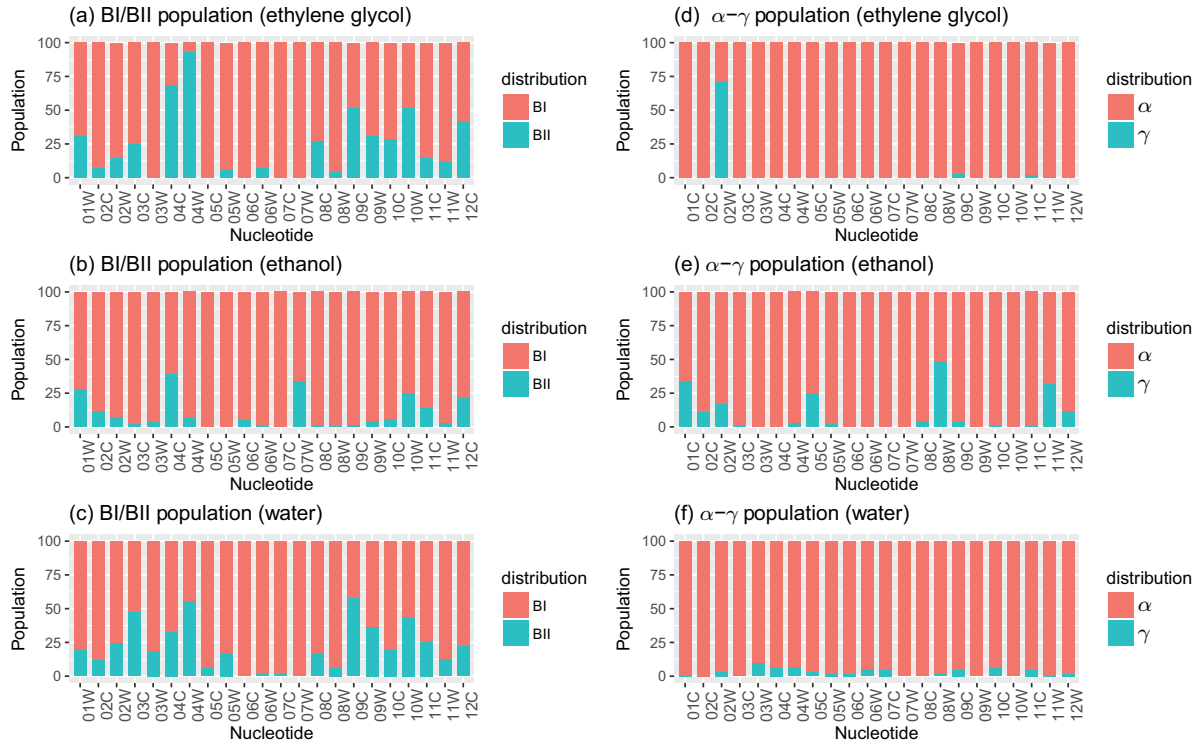


Fig. 10. (color online) Population of BI /BII substate in (a) ethylene glycol, (b) ethanol, and (c) aqueous solutions. Population of canonical α/γ conformation in (d) ethylene glycol, (e) ethanol, and (f) aqueous solutions.

In ethylene glycol solution, there is a non-canonical α/γ torsion at 2G step, the population of canonical $g-/g+$ state is 29%, the population of non-canonical $g-/g-$ (only γ flips) is 35%, $g+/g+$ (only α flips) is 27%, $t/g+$ (only α flips) is 7%, and others is 2%. Flipping tends to occur in one of α/γ, while α/γ has a similar probability of flipping. The probability of α/γ flipping at the same time is only 2%. In ethylene glycol solution, the non-canonical α/γ flips with low probability at other bases. In ethanol solution, non-canonical α/γ flipping appears at 10% of the time, and the distribution is more dispersed in bases. The non-canonical distribution in aqueous solution is also very scattered, but the non-canonical state in pure water solution only appears 3%. The presence of isolated non-canonical α/γ conformers leads to a local decrease in the twist but has little impact in other helical parameters.[39] The unusual α/γ conformation could be a factor favoring recognition by specific proteins, given that the crystal structures of protein–DNA complexes show a significant percentage of such states.

The sugar puckering is divided into four states based on ranges of the sugar pucker phase angle: north puckering (315°–45°), east puckering (45°–135°), south puckering (135°–225°), and west puckering (225°–315°). In ideal B-DNA double helix, the sugar ring is C2'-endo folded belonging to south region. In ideal A-DNA double helix, the sugar ring is C3'-endo folded and belongs to north region. Figure 11 shows population a histogram of population distributions of sugar puckering angles in different solutions. In Fig. 11(a), north–east stands for C4'-exo with an angular range of 36°–72° between the north and east regions, and east–south stands for C1'-exo with a range of angles of 108°–144° between the east and south regions.

As expected for water and ethylene glycol solutions, the sugar puckering is mostly in the south region, accounting for 82% and 70%, respectively. There is also an average of 16% conformation in the east puckering for the aqueous solution, and average of 27% conformation in the east–south mixing zone for the ethylene glycol solution. Among them, the south region (C2'-endo) belongs to the ideal B-DNA, implying that the DNA configuration is a mixture of type B and other configurations (east puckering) in water and glycol solutions. In contrast, the two conformations south and north puckering

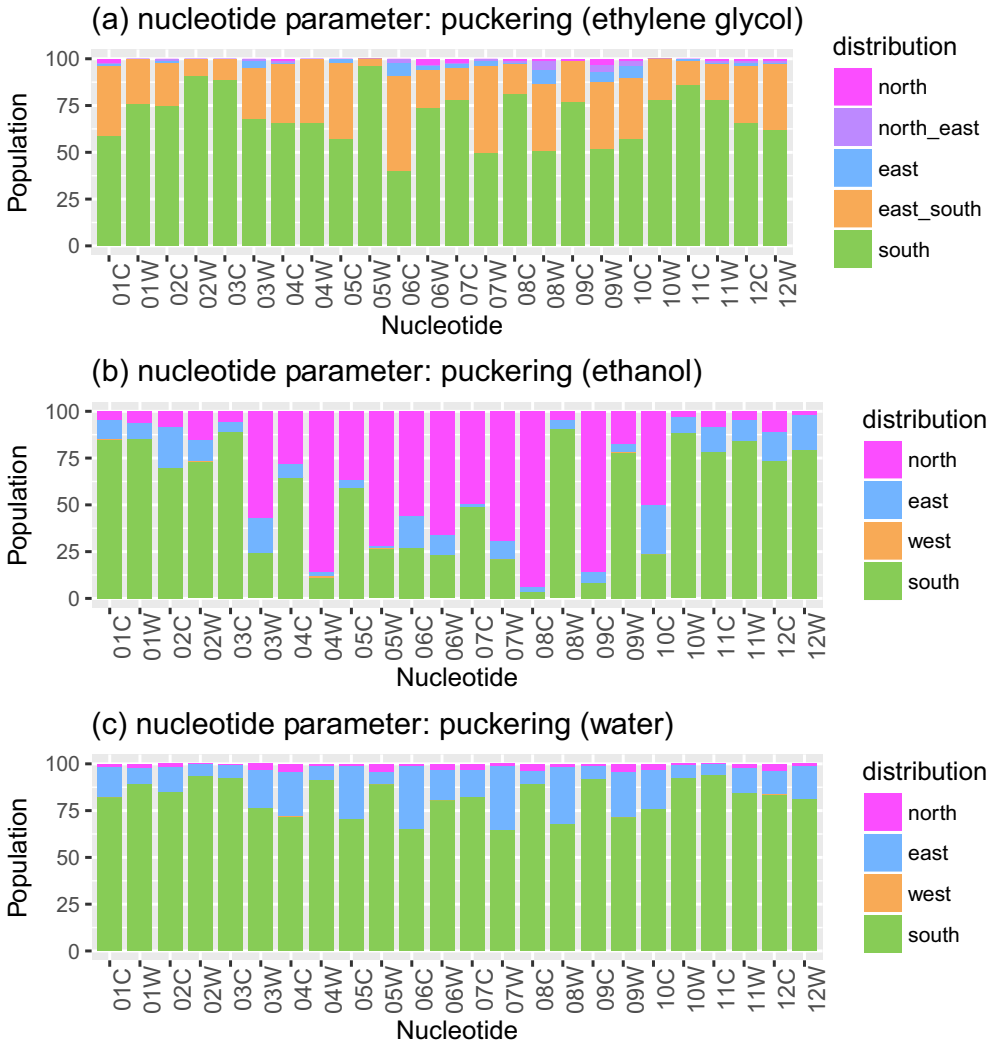


Fig. 11. (color online) Population of nucleotides with north puckering (pink), east puckering (blue), west puckering (yellow), and south puckering (green) in (a) ethylene glycol, (b) ethanol, and (c) aqueous solutions, respectively.

in the ethanol solution are in a dynamic equilibrium, accounting for 55% and 35%, respectively, and the north puckering is mainly in the middle 8 base pairs. The south to north transition produces a moderate increase in the width of the minor groove,[39] among which the north region (C3'-endo) belongs to the ideal A-DNA, which means that the DNA configuration is a mixture of type B and type A in ethanol solutions (10% east puckering).

Study of the DNA backbone torsion shows that the conformations of DNA in ethylene glycol and aqueous solutions are more similar, whereas those in ethanol/water mixtures are clearly different from each other. DNA is more prevalent in B-type conformations in ethylene glycol and aqueous solutions, whereas in ethanol/water mixtures, it is more prevalent in A–B mixed conformations.

To evaluate the DNA structure changes in detail, 8 distinct helical parameters, which are distinguishing between A and B forms, are calculated[52] and presented in Fig. 12. Overall, the conformation of DNA in ethylene glycol and aqueous solution is not very different, but the conformation of DNA in ethanol is quite different, which is consistent with the above conclusion. For ethylene glycol and aqueous solutions, they show a weak positive inclination to the helical axis (inclinations are 5 and 4, respectively) and are moderately shifted towards the major groove (X-displacements are -0.7 and -0.8, respectively). The inter-base pair parameters show an average twist of 34.7°, 34.3° and a roll of 1.6°, 1.5°.

For the base pair helical axis parameters, the standard deviation at the 8T17A base pair is larger in ethanol solution. Table 3 shows the effect of backbone torsional changes on other helical parameters. In ethanol solution, $\mathrm{BI}_{\mathrm{Ethanol}} > \mathrm{BI}_{\mathrm{Water}}$ or $\mathrm{BI}_{\mathrm{Glycol}}$, $g-/g+_{\mathrm{Ethanol}} < g-/g+_{\mathrm{Water}}$ or $g-/g+_{\mathrm{Glycol}}$, $\mathrm{South}_{\mathrm{Ethanol}} < \mathrm{South}_{\mathrm{Water}}$ or $\mathrm{South}_{\mathrm{Glycol}}$. In Fig. 12, this is reflected as the decrease of twist, the increase of roll, and the increase of minor groove width for ethanol solution. In addition, other structural parameters in the ethanol solution also differ greatly from those in the water (ethylene glycol) solution, including of inclination, major groove width, and minor groove depth.

Table 3. The effect of backbone torsional changes on the helical parameters.

Helical parameters	BI to BII	$g-/g+$ to non-canonical	South to north
Twist	↑	↓	–
Roll	↓	–	–
Minor groove width	–	–	↑

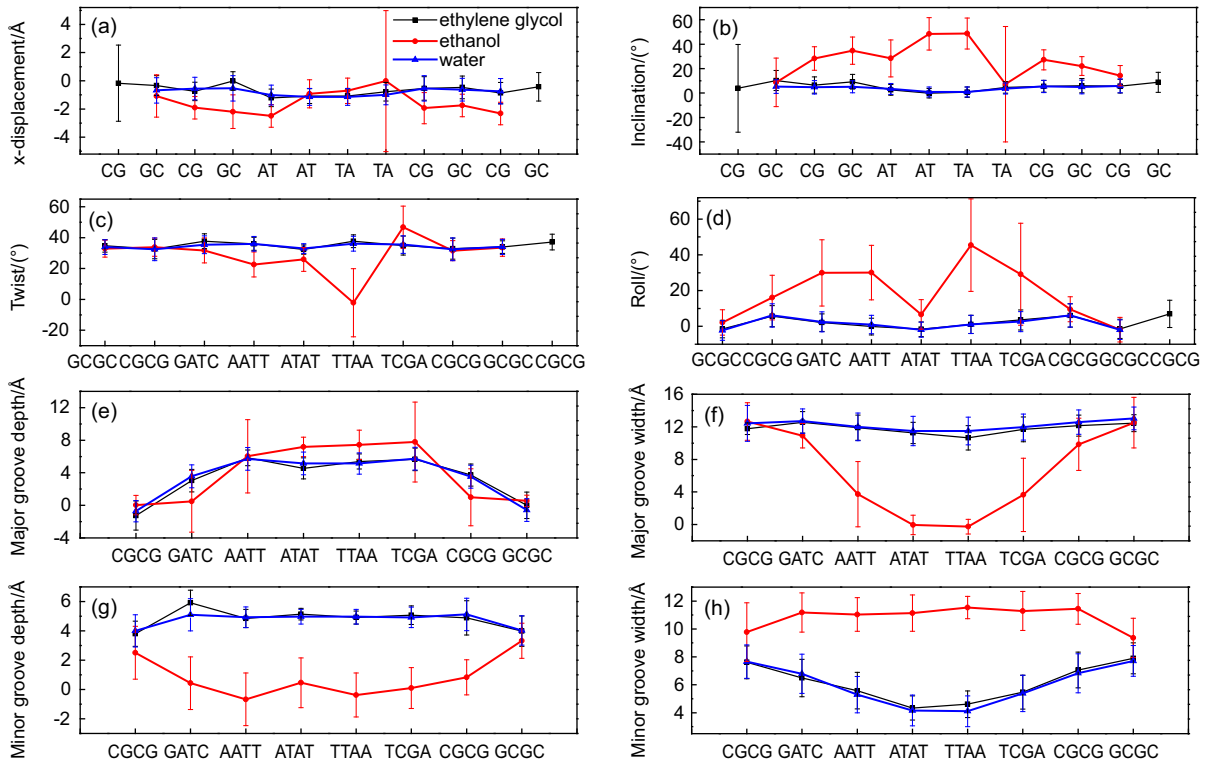


Fig. 12. (color online) Comparison of average values of helical parameters per base pair step from ethylene glycol, ethanol, and aqueous solutions.

4. Conclusion

We have presented the properties of ethylene glycol solutions and the conformational transitions of DNA on the structure in ethylene glycol solutions by using a large scale molecular simulations.

We found that the hydrogen bond network structure changes induced by the concentration increase of the ethylene glycol solutions through different kind of distribution functions such as RDF and SDF. Both of these distribution functions give the picture that the number of molecules in

the first shell layer between ethylene glycol–water and water–water molecules increased with the increase of the concentration of ethylene glycol solutions. On the other hand, the distribution of the first shell molecules between the ethylene glycol–ethylene glycol molecules is not obviously affected by the concentration of the ethylene glycol solutions. Furthermore, the intermolecular hydrogen bond interactions occur between water–water, ethylene glycol–water, and ethylene glycol–ethylene glycol, with the following order of priority: water–water > ethylene glycol–water > ethylene glycol–ethylene glycol.

Under the conditions of 293 K and 1 atm pressure, the conformations of 171d in ethylene glycol solutions are stable and no major distortions are found except for long-life terminal damage, which usually occurs on the terminal cytosine without appearing on the paired guanine. The cytosine undergoes a torsional flip to form a sub-state with a hydrogen-bond forming between the oxygen atom (on the sugar ring of the terminal cytosine) and the hydrogen atom (on the sugar ring of the 3G). In addition, the hydrogen atom (on the sugar ring of the terminal cytosine) formed hydrogen bond with the oxygen atom (on the sugar ring of the 4C).

Analysis of the total number of Watson–Crick hydrogen bonds in ethylene glycol solution shows occurrence of a situation where the number of the hydrogen bonds is higher than the ideal number (32 hydrogen bonds). It may be due to the double-stranded DNA reversibly attaining a more compact and less extended conformation with a smaller solvent shell in ethylene glycol than in water.

The double-stranded DNA in ethylene glycol solution were similar to the structure of DNA in the aqueous solutions but more compact. The reason may be that (i) ethylene glycol molecules and water molecules have the consistent interactional mechanism for the DNA (hydrogen bonds), such that the conformation of DNA in ethylene glycol solutions is similar to that in aqueous solution; (ii) ethylene glycol molecules have higher hydrogen bonding capacity than water molecules to bind to phosphate acid negative oxygen atoms in DNA. The ethylene glycol molecules preferentially distribute around the phosphate groups and minor groove to influence DNA conformation by hydrogen interactions, while water molecules prefer to reside in the grooves. Thus the negatively charged phosphate groups are screened (the repulsion will unwind and extend the helix). In addition, we also added the DNA structure in the ethanol/water mixed solution as a contrast and found that the DNA was in an A–B hybrid state in the ethanol/water mixed solution.

This paper presents a unified quantitative and qualitative method of describing the transitions of DNA conformation in ethylene glycol solutions based on molecular simulation. Although various relevant factors, such as the concentrations of solutions, pressure conditions, and species of ions, were investigated, further studies of these effects and their underlying physical mechanisms are suggested.

References

[1] Chaplin M 2006 *Nat. Rev. Mol. Cell Bio.* **7** 861
[2] Huang Z Q, Yang R Y, Jiang W Z and Zhang Q L 2016 *Chin. Phys. Lett.* **33** 013101
[3] Abolfath R M, van Duin A C T and Brabec T 2011 *J. Phys. Chem. A* **115** 11045
[4] Wang X, Li M, Ye F F and Zhou X 2017 *Acta Phys. Sin.* **66** 150201 (in Chinese)
[5] Cyphers S, Ruff E F, Behr J M, Chodera J D and Levinson N M 2017 *Nat. Chem. Biol.* **13** 402
[6] Ulrich K, Galvosas P, Kärger J and Grinberg F 2009 *Phys. Rev. Lett.* **102** 037801
[7] Fernández A 2014 *J. Chem. Phys.* **140** 221102
[8] Lynden-Bell R M and Debenedetti P G 2005 *J. Phys. Chem. B* **109** 6527
[9] Ball P 2005 *Nature* **436** 1084
[10] Gu B, Zhang F S, Wang Z P and Zhou H Y 2008 *Phys. Rev. Lett.* **100** 088104
[11] Errington J R and Debenedetti P G 2001 *Nature* **409** 318
[12] Chakraborty D and Wales D J 2017 *Phys. Chem. Chem. Phys.* **19** 878
[13] Samanta S, Mukherjee S, Chakrabarti J and Bhattacharyya D 2009 *J. Chem. Phys.* **130** 115103
[14] Banavali N K and Roux B 2005 *J. Am. Chem. Soc.* **127** 6866
[15] Lange A W and Herbert J M 2009 *J. Am. Chem. Soc.* **131** 3913
[16] Shen X, Atamas N A and Zhang F S 2012 *Phys. Rev. E* **85** 051913
[17] Maaloum M, Beker A F and Muller P 2011 *Phys. Rev. E* **83** 031903
[18] Wang Y W and Yang G C 2017 *Chin. Phys. B* **26** 128706
[19] Lindahl T 2016 *Nat. Rev. Mol. Cell Bio.* **17** 335
[20] Eliasson R, Hammarsten E and Lindahl T 1962 *Biotechnol. Bioeng.* **4** 53
[21] Anagnostopoulos C and Spizizen J 1961 *J. Bacteriol.* **81** 741
[22] Pogozelski W K and Tullius T D 1998 *Chem. Rev.* **98** 1089
[23] Tullius T D and Greenbaum J A 2005 *Curr. Opin. Chem. Biol.* **9** 127
[24] Aydogan B, Bolch W E, Swarts S G, Turner J E and David T 2008 *Radiat. Res.* **169** 223
[25] Friedland W, Jacob P, Paretzke H G, Merzagora M and Ottolenghi A 1999 *Radiat. Environ. Biophys.* **38** 39
[26] Abolfath R M, Biswas P K, Rajnarayanam R, Brabec T, Kodym R and Papiez L 2012 *J. Phys. Chem. A* **116** 3940
[27] Merzel F, Fontaine-Vive F, Johnson M R and Kearley G J 2007 *Phys. Rev. E* **76** 031917
[28] Smirnov D A, Morley M, Shin E, Spielman R S and Cheung V G 2009 *Nature* **459** 587
[29] Rak J, Chomicz L, Wiczk J, Westphal K, Zdrowowicz M, Wityk P, Zyndul M, Makurat S and Golon Ł 2015 *J. Phys. Chem. B* **119** 8227
[30] Hamelberg D, McFail-Isom L, Williams L D and Wilson W D 2000 *J. Am. Chem. Soc.* **122** 10513
[31] Korolev N, Lyubartsev A P, Laaksonen A and Nordenskiold L 2003 *Nucleic Acids Res.* **31** 5971
[32] Noy A, Pérez A, Laughton C A and Orozco M 2007 *Nucleic Acids Res.* **35** 3330
[33] Pastor N 2005 *Biophys. J.* **88** 3262
[34] Freddolino P L, Arkhipov A S, Larson S B, McPherson A and Schulten K 2006 *Structure* **14** 437
[35] Shen H, Cheng W and Zhang F S 2015 *RSC Adv.* **5** 9627
[36] Martínez L, Andrade R, Birgin E G and Martínez J M 2009 *J. Comput. Chem.* **30** 2157
[37] Schweitzer B I, Mikita T, Kellogg G W, Gardner K H and Beardsley G P 1994 *Biochemistry* **33** 11460
[38] Dans P D, Danilāne L, Ivani I, Dršata T, Lankaš F, Hospital A, Walther J, Pujagut R I, Battistini F, Gelpí J L, Lavery R and Orozco M 2016 *Nucleic Acids Res.* **44** 4052
[39] Pérez A, Luque F J and Orozco M 2007 *J. Am. Chem. Soc.* **129** 14739
[40] Dršata T, Pérez A, Orozco M, Morozov A V, Šponer J and Lankaš F 2013 *J. Chem Theory Comput.* **9** 707
[41] Mark P and Nilsson L 2002 *J. Phys. Chem. B* **106** 9440

[42] Ivani I, Dans P D, Noy A, Pérez A, Faustino I, Hospital A, Walther J, Andrio P, Goñi R, Balaceanu A, Portella G, Battistini F, Gelpí J L, González C, Vendruscolo M, Laughton C A, Harris S A, Case D A and Orozco M 2016 *Nat. Methods* **13** 55
[43] Darden T, York D and Pedersen L 1998 *J. Chem. Phys.* **98** 10089
[44] Allen M P and Tildesley D J 1989 *Computer Simulation of Liquids* (Oxford University Press)
[45] Hess B, Bekker H, Berendsen H J C and Fraaije J G E M 1997 *J. Comput. Chem.* **18** 1463
[46] Berendsen H J C, Grigera J R and Straatsma T P 1987 *J. Phys. Chem.* **91** 6269
[47] Zhang F S and Lynden-Bell R M 2005 *Phy. Rev. E* **71** 021502
[48] Lindahl E, Hess B and van der Spoel D 2001 *J. Mol. Model.* **7** 306
[49] Tsierkezos N G and Molinou I E 1998 *J. Chem. Eng. Data* **43** 989
[50] Nose S and Klein M L 1983 *Mol. Phys.* **50** 1055
[51] Berendsen H J C, Postma J P M, van Gunsteren W F, DiNola A and Haak J R 1984 *J. Chem. Phys.* **81** 3684
[52] Lavery R, Moakher M, Maddocks J H, Petkeviciute D and Zakrzewska K 2009 *Nucleic Acids Res.* **37** 5917
[53] Lavery R and Sklenar H 1989 *J. Biomol. Struct. Dyn.* **6** 655
[54] Kumar R, Schmidt J R and Skinner J L 2007 *J. Chem. Phys.* **126** 204107
[55] Herskovits T T 1962 *Arch. Biochem. Biophys.* **97** 474
[56] Zgarbová M, Otyepka M, Šponer J, Lankaš F and Jurečka P 2014 *J. Chem. Theory Comput.* **10** 3177
[57] Leroy J L, Kochoyan M, Huynh-Dinh T and Guéron M 1988 *J. Mol. Biol.* **200** 223
[58] Andreatta D, Sen S, Lustres J L P, Kovalenko S A, Ernsting N P, Murphy C J, Coleman R S and Berg M A 2006 *J. Am. Chem. Soc.* **128** 6885
[59] McConnell K J and Beveridge D L 2000 *J. Mol. Biol.* **304** 803
[60] Dickerson R E, Drew H R, Conner B N, Wing R M, Fratini A V and Kopka M L 1982 *Science* **216** 475
[61] Allahyarov E, Gompper G and Lowen H 2013 *Phys. Rev. E* **69** 041904
[62] Saenger W, Hunter W N and Kennard O 1986 *Nature* **324** 385
[63] Hartmann B, Piazzola D and Lavery R 1993 *Nucleic Acids Res.* **21** 561
[64] Schwieters C D and Clore G M 2007 *Biochemistry* **46** 1152
[65] Djuranovic D and Hartmann B 2004 *Biopolymers* **73** 356
[66] Heddi B, Oguey C, Lavelle C, Foloppe N and Hartmann B 2010 *Nucleic Acids Res.* **38** 1034
[67] Wecker K, Bonnet M C and Meurs E F 2002 *Nucleic Acids Res.* **30** 4452
[68] Madhumalar A and Bansal M 2005 *J. Biomol. Struct. Dyn.* **23** 13
[69] Dans P D, Faustino I, Battistini F, Zakrzewska K, Lavery R and Orozco M 2014 *Nucleic Acids Res.* **42** 11304