Development of coarse-graining DNA models for single-nucleotide resolution analysis

BY KENTARO DOI*, TOMOAKI HAGA, HIROFUMI SHINTAKU AND SATOYUKI KAWANO

Department of Mechanical Science and Bioengineering, Graduate School of Engineering Science, Osaka University, Osaka 560-8531, Japan

Recently, analytical techniques have been developed for detecting single-nucleotide polymorphisms in DNA sequences. Improvements of the sequence identification techniques has attracted much attention in several fields. However, there are many things that have not been clarified about DNA. In the present study, we have developed a coarse-graining DNA model with single-nucleotide resolution, in which potential functions for hydrogen bonds and the π-stack effect are taken into account. Using Langevin-dynamics simulations, several characteristics of the coarse-grained DNA have been clarified. The validity of the present model has been confirmed, compared with other experimental and computational results. In particular, the melting temperature and persistence length are in good agreement with the experimental results for a wide range of salt concentrations.

Keywords: DNA; coarse graining; melting; persistence length; Langevin dynamics

1. Introduction

Rapid development of DNA sequencing techniques make it possible to find single-nucleotide polymorphisms (SNPs) that cause serious diseases such as cancer (Miki et al. 1994; Risch et al. 2001; Velasco et al. 2007). For example, denaturing gradient gel electrophoresis (DGGE) is known to be one of the major techniques that can detect SNPs in several hundred base alignments (Myers et al. 1985; Sheffield et al. 1989; Abrams et al. 1995; Michikawa et al. 1997). DNA fragments that have SNPs in the sequences show different electrophoretic flows caused by different interactions between dissociated base pairs and the surrounding gel media (Ganguly et al. 1993). However, the detailed mechanism of mobility reduction is not yet understood. There are some difficulties in clarifying the nature because the target, about 50nm in length, is very small for experimental observation, but very large for all atom molecular simulations.

In order to overcome these difficulties, several coarse-graining methods have been developed. DNA models such as the bead-spring model (Deutsch 1988; Chen et al. 2005; Streek et al. 2005; Nagahiro et al. 2007; Trahan & Doyle 2009), the worm-like chain (WLC) model (Jian & Vologodskii 1997; Wang & Gao 2005) and

*Author for correspondence (doi@me.es.osaka-u.ac.jp).

One contribution of 13 to a Theme Issue ‘The virtual physiological human: computer simulation for integrative biomedicine I’.
the double-strand model (Cocco & Monasson 1999; Lyubartsev & Laaksonen 1999; Drukker et al. 2001; Tepper & Voth 2005; Knotts et al. 2007; Mielke et al. 2008) are well known. In the bead-spring model, the coarse-grained particles are about 100 nm in diameter, and the particles are connected with springs. Hydrodynamic interactions between particles can be treated effectively, although bending elasticity of DNA is not considered. Bead-spring models are available for the case that the contour length $L$ is larger than the persistence length $P$. It is known that flow in very narrow or confined channels is effectively simulated using the bead-spring models (Chen et al. 2005). On the other hand, the WLC model is known to be effective when $L$ is nearly equal to $P$ (Jian & Vologodskii 1997; Wang & Gao 2005). A potential function for bond angles is considered based on the bending elasticity.

In order to simulate further characteristics of DNA, such as melting, hybridization or loop structures, the double-stranded structure of DNA is required to be constructed in the coarse-graining procedure. These characteristics are caused by making or breaking hydrogen bonds between base molecules. In particular, differences between adenine-thymine (AT) and guanine-cytosine (GC) base pairs (bps) are the nature of the phenomena. Several kinds of the double-strand model have been developed for each purpose. For example, Drukker et al. (2001) investigated the melting property of DNA using a model in which a single nucleotide was replaced by two particles. The coarse-grained particles, which was 1 nm in diameter, represented parts of a base and backbone. They succeeded in realizing the melting phenomenon of AT and GC fragments. Knotts et al. (2007) replaced a single nucleotide with three particles that represented a phosphate group, sugar and a base. In addition, negative charges on phosphate groups were considered and the distribution of counter cations was treated using the Debye–Hückel approximation. Using the model of 60 bps, they found that not only were loop structures observed in AT-rich parts, but also the hybridization phenomenon. Mielke et al. (2008) constructed a model in which three nucleotides were replaced by a particle. Their model succeeded in realizing the persistence length of double-stranded chains and clarified that counter cations around DNA fragments obeyed the Poisson–Boltzmann distribution. However, their model did not have single-nucleotide resolution.

In the present study, we develop a coarse-graining model focusing on the dynamical aspects of double-stranded B–DNA. One of our interests is to clarify the detecting process of different DNA sequences in electrophoretic flow. Therefore, our goal is to make this model available for fluid-dynamics simulations of DNA, which are influenced by viscosity, salt concentrations and external electric fields. The simulation techniques have not been developed sufficiently, although each property of DNA was realized by coarse-graining methods from previous work. In our model, a single nucleotide is replaced by a coarse-grained particle. Several kinds of potential functions are adopted for interactions between the particles. We define hydrogen bonds in base pairs by Morse-type functions and particularly adopt weak harmonic oscillators for the $\pi$-stack effect in the base-pair stacks. Using the present model, the melting temperature and persistence length of DNA sequences are evaluated; these are variables for measuring the bending characteristics. The relation between persistence length and salt concentrations is discussed and compared with experimental data. The scalability of the model from 10 to 155 bps is demonstrated.

Phil. Trans. R. Soc. A (2010)
Development of coarse-graining DNA models

2. Development of the coarse-graining DNA model

(a) Potential energy functions in coarse-grained DNA

An SNP is observed within about 1000 bps. Single-base-pair differences can be found clearly as a result of electrophoresis. However, the behaviour of DNA fragments in the gel solution has not been understood sufficiently. In the present study, we develop a coarse-graining model of DNA that has single-nucleotide resolution. A schematic illustration of our model is presented in figure 1. Figure 1a shows an image of the all-atom model of double-stranded DNA and (b) represents its coarse-grained model. The coarse-grained particles are located at positions of each P atom in the backbones. Interactions between the particles are represented by several potential functions,

\[ V_{\text{total}} = V_{\text{bond}} + V_{\text{angle}} + V_{\text{bp}} + V_{\text{stack}} + V_{\text{elec}}. \] (2.1)

Here, bond lengths and bond angles are assumed to be oscillated around the equilibrium structures,

\[ V_{\text{bond}} = \sum_{i=1}^{N_{\text{bond}}} \frac{1}{2} k_\text{b}(d_i - d_{\text{eq}})^2 \] (2.2)

and

\[ V_{\text{angle}} = \sum_{i=1}^{N_{\text{angle}}} \frac{1}{2} k_\theta(\theta_i - \theta_{\text{eq}})^2. \] (2.3)

Figure 1. Schematic illustrations of B–DNA: (a) all-atom model and (b) coarse-graining model in which positions of particles correspond to those of P atoms in DNA.
These potentials can be well applied to many models in which the interactions can be assumed to be harmonic oscillations. Here, \( k_b \) and \( k_q \) are force constants in the oscillators and \( d_{eq} \) and \( \theta_{eq} \) are bond length and bond angle in the equilibrium structures, respectively. The \( V^{\text{bond}} \) and \( V^{\text{angle}} \) depend on chemical bonds and molecular structures, which are dominated by short-range interactions. They are defined between the nearest-neighbour particles that represent single nucleotides. On the other hand, the long-range interactions are represented by \( V^{\text{elec}} \). Moreover, hydrogen bonds between base molecules are known to be written by Morse-type functions as follows:

\[
V^{\text{bp}} = \frac{N_{\text{bp}}}{\sum_{i=1}^{N_{\text{bp}}} D \left[ \exp \left( -\alpha (l_i - l_{eq}) \right) - 1 \right]^2 - D}.
\] (2.4)

A binding energy of hydrogen bonds given by \( D \) in equation (2.4) is different between each complementary base pair. Hydrogen bonds in a GC pair are considered to be 1.5 times stronger than those in an AT pair because of the difference in the number of hydrogen bonds. The constant \( l_{eq} \) is an equilibrium distance between base pairs and \( \alpha \) is a parameter depending on the force constants. Furthermore, the potential function caused by \( \pi-\pi \) interactions in base-pair stacks is defined as follows:

\[
V^{\text{stack}} = \frac{N_{\text{stack}}}{\sum_{i=1}^{N_{\text{stack}}} \frac{1}{2} k_s (h_i - h_{eq})^2}.
\] (2.5)

The interaction between base-pair stacks is assumed to be that of harmonic oscillations. The constants \( k_s \) and \( h_{eq} \) are the force constant and the equilibrium distance between the base-pair stacks, respectively. The \( h_{eq} \) is estimated to be 3.40 Å and referred to as the double-stranded structure of B–DNA. The force constant of \( k_s \) is represented by \( k_s = k_B T / \gamma^2 \) and the averaged thermal fluctuation is assumed to be \( \gamma = 0.10 h_{eq} \). This stacking potential plays an important role in the double-stranded structures of DNA. Furthermore, negative charges on the phosphate groups are taken into account. It is known that the electrophoresis of DNA fragments is caused by the response of these negative charges to external electric fields. The electrostatic interaction is represented by means of the Debye–Hückel approximation,

\[
V^{\text{elec}} = \sum_{i<j}^N \frac{q_i q_j}{4 \pi \varepsilon_0 \varepsilon_k r_{ij}} e^{-r_{ij}/\lambda_D} \quad \text{and} \quad \lambda_D = \left( \frac{\varepsilon_0 \varepsilon_k k_B T}{2 N_A e^2 I} \right)^{1/2}.
\] (2.6)

Charges of \( q_i \) and \( q_j \) are located at the distance \( r_{ij} \). \( N \) is the number of charged particles. The constants \( \varepsilon_0 \) and \( \varepsilon_k \) are the permittivity of the vacuum and the relative permittivity of the aqueous solution, respectively. The factor \( \lambda_D \) is the Debye length. \( N_A \), \( k_B \) and \( T \) are Avogadro’s number, the Boltzmann constant and temperature, respectively. \( I \) is the ionic strength of the salt. In the present study, negative charges on each particle are fixed at \(-1\) in the elementary charge.

The characteristics of the potential functions are verified, compared with the melting temperature of DNA fragments. In order to recognize the result, parameters in each potential function are determined as below. The force constant
Development of coarse-graining DNA models

of a chemical bond is defined as $k_b = k_B T / \delta^2$. The $\delta$ represents an average of the thermal fluctuation and the value is set to be 0.01 $d_{eq}$. The bending elasticity of the backbone is represented by the angle potential of equation (2.3). The force constant of the angle potential is defined as $k_b = k_B T P_{ss} / d_{eq}$ (Hagerman & Zimm 1981). Here, $P_{ss}$ is the persistence length of a single-stranded DNA. The value is assumed to be 4.0 nm (Tinland et al. 1997). The equilibrium angle of $\theta_{eq}$ is set to be 180.0° based on an assumption of the isotropic nature of the bending elasticity.

Figure 2 shows the angles between the nearest neighbours in the optimized base-pair stacks. The angles are estimated from the inner product of the base-pair axes between the nearest neighbours. The plots show the averages of 10 bp fragments. According to $\gamma$, the stacking angle settles at different values. When $\gamma$ is 0.10 $h_{eq}$, the averaged angle between the stacking pair converges to 35.4°. The Morse potentials for hydrogen bonds are determined reasonably in order to realize the melting temperature. In equation (2.4), the binding energies of hydrogen bonds $D_{GC}$ and $D_{AT}$ are estimated as 28.8 and 19.2 kJ mol$^{-1}$, respectively. Moreover, the parameter $\alpha$ in equation (2.4) influences the force constant defined as $2\alpha^2 D$. The force constant is a key factor to determine the melting temperature of the coarse-graining model. The constants, $\alpha_{GC}$ and $\alpha_{AT}$, are set to be 4.5 Å and 4.0 Å, respectively. Figure 3 shows the computational results of the melting temperature for 10 bp poly(dG) · poly(dC) and 10 bp poly(dA) · poly(dT). These plots were obtained from the averages of 200 ns samplings. The validity of the values are confirmed to be in agreement with the results of the nearest-neighbour method (Santalucia et al. 1996). The melting temperature is determined from half-cleaved structures of the base-pair fragments. A base-pair distance longer than $l_{eq} + 6 / \alpha$ is considered to be cleaved according to the consideration of Cocco & Monasson (1999). The temperatures are estimated as 338 K and 304 K for poly(dG) · poly(dC) and poly(dA) · poly(dT), respectively. On the other hand, the nearest-neighbour method estimates the melting temperatures as 330 K and 295 K for poly(dG) · poly(dC) and
poly(dA) · poly(dT), respectively. The present results are in good agreement with
the results from the nearest-neighbour method within the errors of 2.4 per cent
for poly(dG) · poly(dC) and 3.0 per cent for poly(dA) · poly(dT). In the present
study, using the potential functions and parameters described above, the melting
temperatures and bending properties of the coarse-grained DNA are computed
and compared with other theoretical and experimental results.

(b) Langevin-dynamics simulation

Langevin-dynamics computations are carried out for DNA fragments in an
aqueous solution. The equation of motion of the $i$th particle that moves with
velocity $v_i$ is written as

$$m_i \ddot{v}_i = F^d_i - \nabla V^{\text{total}} + F^B_i,$$

where $m_i$ is the mass, $F^d_i$ is the viscous drag and $F^B_i$ is the random force caused
by thermal fluctuations. The inertial motion of the coarse-grained particles is so
small when compared with the other terms that it can be ignored. The random
force is assumed to be generated with the following conditions:

$$\langle F^B_i(t) \rangle = 0 \quad (2.8)$$

and

$$\langle F^B_i(t) \cdot F^B_j(t') \rangle = 6k_B T \zeta \delta_{ij} \delta(t - t'). \quad (2.9)$$

The viscous drag is defined as $F^d_i = -\zeta v_i$ according to Stokes’s law and the friction
coefficient $\zeta$ can be replaced by $6\pi \eta a$, with viscosity of water $\eta$ and radius of
the particle $a$. In the present computation, the radius of a single nucleotide is
represented by $a$. The stochastic Runge–Kutta procedure (Honeycutt 1992) is
employed to integrate the Langevin equation. A time step of 10 fs is chosen for
the numerical integrations.
Development of coarse-graining DNA models

Figure 4. Degrees of freedom in base stacks: (a) a model by El Hassan & Calladine (1995) and (b) the present model.

Table 1. Averaged parameters of the degrees of freedom in coarse-grained DNA.

<table>
<thead>
<tr>
<th></th>
<th>1 mM</th>
<th>10 mM</th>
<th>25 mM</th>
<th>50 mM</th>
<th>100 mM</th>
<th>150 mM</th>
<th>MD a</th>
<th>crystal b</th>
</tr>
</thead>
<tbody>
<tr>
<td>⟨twist⟩ (°)</td>
<td>34.0</td>
<td>34.8</td>
<td>35.1</td>
<td>35.4</td>
<td>35.6</td>
<td>35.7</td>
<td>33.1</td>
<td>35.4</td>
</tr>
<tr>
<td>⟨slide⟩ (Å)</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.01</td>
<td>−0.32</td>
<td>0.35</td>
</tr>
<tr>
<td>⟨shift⟩ (Å)</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>⟨rise⟩ (Å)</td>
<td>3.81</td>
<td>3.63</td>
<td>3.54</td>
<td>3.47</td>
<td>3.41</td>
<td>3.37</td>
<td>3.33</td>
<td>3.32</td>
</tr>
</tbody>
</table>

a Referred to Šponer & Lankaš (2006).
b Referred to Olson et al. (1998).

3. Results and discussion

(a) Validation of coarse-graining DNA model

Taking the thermal fluctuations into account, Langevin-dynamics simulations have been carried out to analyse the dynamical properties of DNA fragments. According to the topological symmetry of B–DNA, El Hassan & Calladine (1995) suggested several indices to characterize molecular condition. In their model, a set of two base pairs was divided into six degrees of freedom, such as tilt, roll, twist, slide, shift and rise. Figure 4a shows a schematic of six degrees of freedom. Base molecules represented by plates are connected rigidly with another complementary pair. Three axes are defined for translations of the base pair and the other three are for rotations. In our model shown in figure 4b, however, two nucleotides replaced by particles are connected with a line, and only four degrees of freedom, such as twist, slide, shift and rise, can be compared with the model of figure 4a. The simulations for a 155 bp sequence were carried out for 100 ns at the temperature of 300 K. Table 1 shows each parameter resulting from Langevin-dynamics analysis for several salt concentrations. The time-averaged rotational angle ⟨twist⟩ increases as a function of the salt concentration. This is because increase of the salt concentration causes screening of negative charges on the backbone, the electrostatic repulsion between the particles is weakened and the force from the stacking potential turns to be effective to twist the double strand. On the other hand, ⟨rise⟩ decreases and the molecular contour length is shortened as ⟨twist⟩ increases. In addition, ⟨slide⟩ and ⟨shift⟩ are not influenced more by the screening effect. The results are compared with other molecular dynamics computations (Šponer & Lankaš 2006) and crystal-structure analyses.
Figure 5. Temperature dependency on base-pair dissociation ratios of several sequences. Each plot is an average of 100 ns samplings in Langevin-dynamics simulations. Melting temperatures are determined from the half-cleaved points. Solid line with circles, L19AS2; dashed line with empty squares, L33B9; dotted line with triangles, L42B18; dotted line with filled circles, L48AS; dashed-dotted line with filled squares, L60B36.

Table 2. Sequences used for analysing melting temperature.

<table>
<thead>
<tr>
<th>type</th>
<th>sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>L19AS2</td>
<td>5’-ATA ATA AAG GCG GTC CGC C-3’</td>
</tr>
<tr>
<td>L33B9</td>
<td>5’-CCG CCC GCG GCC TTT ACT AAA GGC CGC TGC GCC-3’</td>
</tr>
<tr>
<td>L42B18</td>
<td>5’-CCG CCC GCG GCC TTA ATA CTT AAG TAT TAT GGC CGC TGC GCC-3’</td>
</tr>
<tr>
<td>L48AS</td>
<td>5’-CAT AAT ACT TTA TAT TTA ATT GGC GGC GCA CGG GAC CCG TGC GCC GCC-3’</td>
</tr>
<tr>
<td>L60B36</td>
<td>5’-CCG CCC GCG GCC TTA TTA CAT TTA ATT CTT AAG TAT TAT AAG TAA TAT GCC CGC TGC GCC-3’</td>
</tr>
</tbody>
</table>

(Olson et al. 1998). Our results show good agreement with those theoretical and experimental analyses. Furthermore, our averaged parameters show the dependency on salt concentrations. Therefore, the coarse-grained model is expected to realize dynamics of DNA fragments in some kinds of aqueous solution.

**(b) Melting temperature of DNA fragments**

In order to discuss the melting properties of our model, the melting temperature was calculated using the Langevin-dynamics simulation. Several B–DNA sequences presented in table 2 were prepared for comparison with experimental results (Zeng et al. 2004). As a result of 100 ns samplings in each single trial, ratios of hydrogen-bond breaking were estimated as shown in figure 5. In the present case, the ionic strength of the solvent was set to be 50 mM. The dissociation of a complementary base pair was judged from the bond length of $l_{eq} + 6/\alpha$ (Cocco & Monasson 1999), where $l_{eq}$ and $\alpha$ are parameters of equation (2.4). The melting temperatures, which were determined...
Development of coarse-graining DNA models

Table 3. Melting temperatures at 50 mM in salt concentration.

<table>
<thead>
<tr>
<th>melting temperature (K)</th>
<th>present study</th>
<th>experimenta</th>
<th>error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>L19AS2</td>
<td>333</td>
<td>321</td>
<td>3.7</td>
</tr>
<tr>
<td>L33B9</td>
<td>345</td>
<td>344</td>
<td>0.3</td>
</tr>
<tr>
<td>L42B18</td>
<td>349</td>
<td>339</td>
<td>3.0</td>
</tr>
<tr>
<td>L48AS</td>
<td>349</td>
<td>341</td>
<td>2.3</td>
</tr>
<tr>
<td>L60B36</td>
<td>360</td>
<td>336</td>
<td>7.1</td>
</tr>
<tr>
<td>10 bp poly(dG) · poly(dC)</td>
<td>338</td>
<td>330</td>
<td>2.4</td>
</tr>
<tr>
<td>10 bp poly(dA) · poly(dT)</td>
<td>304</td>
<td>295</td>
<td>3.0</td>
</tr>
</tbody>
</table>

aReferred to Zeng et al. (2004).

from the half-cleaved point, are summarized in table 3. The half-cleaved points were determined from the point at the intersection with 0.5 using linear interpolation. In the present model, hydrogen-bond breaking is mainly caused by the $V^{bp}$ adjusted to the melting properties of 10 bp poly(dG) · poly(dC) and poly(dA) · poly(dT). Results from our simulations are usually in agreement with the experimental data (Zeng et al. 2004), particularly for the cases of L19AS2, L33B9, L42B18 and L48AS. The errors from these models are as small as the cases of 10 bp poly(dG) · poly(dC) and poly(dA) · poly(dT). L19AS2 and L48AS sequences have many AT pairs near the edge. Therefore, base-pair cleavages occurred and expanded from the edge as the temperature increased. On the other hand, in L33B9 and L42B18 sequences, GC pairs are concentrated at both ends. These alignments seem to make the structures toughen and retard their melting. It is found that the adjustment for small DNA fragments, such as 10 bp poly(dG) · poly(dC) and poly(dA) · poly(dT), is available for larger models. This coarse-graining technique has scalability about the size and can realize melting characteristics of double-stranded DNA. Meanwhile, the disagreement is remarkable in the case of L60B36, although the error is at most 7.1 per cent. The melting property is known to be enhanced by the expansion of cleavages at the ends or middle parts of a fragment (Barbi et al. 1999; Hwa et al. 2003; Kalosakas et al. 2004; Ares et al. 2005). In the present computations, the dissociation of complementary base pairs tended to expand from the edges at temperatures higher than 350 K and from the middle parts at less than 340 K. Because of these different melting characteristics, the melting temperature of L60B36 was overestimated, compared with the experimental result (Zeng et al. 2004). In the experiment, the L60B36 sequence is considered to have loop structures at the middle parts. Therefore, the melting temperature is expected to be lower than that of the completely closed structure. The present result suggests that loop structures formed at the middle parts of a sequence enhance the melting at lower temperature than cleavages from the ends of GC clamps.

(e) Persistence length

The persistence length of DNA fragments is analysed quantitatively in order to discuss the bending properties. The persistence length is mainly characterized by bending elasticity and Coulomb interaction between the coarse-grained particles.
Baumann et al. (1997) calculated the persistence length by fitting their tension test for DNA fragments to the WLC model. They reported that their results agree well with computational results based on Poisson–Boltzmann theory. The persistence length of our DNA model is analysed using an approximation by Hagerman & Zimm (1981),

\[ R_a = 1.0120 - 0.24813X + 0.0333703X^2 - 0.0019177X^3. \]  

(3.1)

Here, \( R_a = \tau_a / \tau_B \) denotes a proportion of the approximated rotational relaxation time \( \tau_a \) to Broersma’s rotational relaxation time, \( \tau_B \). The variable \( X = L/P \) is defined as a proportion of the contour length \( L \) to the persistence length \( P \) that should be determined. Equation (3.1) is known to be effective for the condition \( L/P \geq 0.5 \) (Hagerman & Zimm 1981). The approximated rotational relaxation time \( \tau_a \) is defined as

\[
\frac{1}{\tau_a} = 3(\langle D_x \rangle + \langle D_y \rangle),
\]  

(3.2)

where \( \langle D_x \rangle \) and \( \langle D_y \rangle \) are time averages of the rotational diffusion coefficient for the \( x \)- and \( y \)-directions, respectively. Broersma’s relaxation time is calculated according to the relation \( 1/\tau_B = 6D_B \). The rotational diffusion coefficient \( D_B \) is defined as follows (Hagerman & Zimm 1981):

\[
D_B = \left( \frac{3k_B T}{\pi \eta h^3} \right) \left[ \ln \left( \frac{h}{b} \right) - 1.57 + 7 \left( \frac{1}{\ln(h/b)} - 0.28 \right)^2 \right],
\]  

(3.3)

where \( h \) denotes the height of the coarse-grained DNA referred to as Broersma’s cylinder and \( b \) is its radius. The value \( h \) is estimated as \( h = n \langle \text{rise} \rangle \), where \( n \) is the number of base pairs. The value \( b \) is determined as the cylindrical radius that denotes the equivalent volume as the coarse-grained base pair,

\[
2 \times \frac{4}{3} \pi a^3 = \pi b^2 \times \langle \text{rise} \rangle.
\]  

(3.4)

In this study, the hydrodynamic radius \( a \) of a single nucleotide is set to be 10 Å. The Langevin-dynamics simulations were performed for a 155 bp fragment at 300 K for 100 ns. Figure 6 shows the molecular structures at 100 ns, obtained in salt concentrations of 1, 10 and 150 mM. The molecular structures depending on the salt concentration are visually characterized. The higher the salt concentration, the longer the distance between both ends are because of the screening effect on the negative charges of the backbones. For qualitative validation, these molecular structures were analysed from the averages of 10 000 time-samplings for 100 ns. Figure 7 shows the salt-concentration dependency of the rotational relaxation time, comparing between Broersma’s model and our coarse-graining model. In both models, the rotational relaxation times tend to be reduced as the salt concentration increases. At each salt concentration, the rotational diffusion coefficients are confirmed to be converged enough during 100 ns samplings, although the figure is omitted here. Substituting the proportion of these two rotational relaxation times in equation (3.1), the relation between the persistence length and the salt concentration can be evaluated. Figure 8 shows a comparison between our computational results and other experimental results (Baumann et al. 1997). The experimental data were applied to some coarse-graining models, such as the inextensible WLC, strong stretching limit.
Development of coarse-graining DNA models

Figure 6. Molecular structures of a 155 bp fragment at 100 ns, depending on salt concentrations: (a) 1 mM; (b) 10 mM; and (c) 150 mM. The higher the salt concentrations, the more the DNA fragments bend in the solvent because of the screening effect on the negative charges.

Figure 7. Rotational relaxation times of a 155 bp fragment as a function of salt concentrations, which influences the persistence length according to equation (3.5). Solid line with open circles, Broersma model; dashed line with filled circles, coarse-grained model.

and extensible WLC. Another estimation using Poisson–Boltzmann theory is also presented in figure 8. In this case, DNA fragments are assumed to have a cylindrical structure that has a homogeneous electronic distribution. The persistence length $l_p$ was estimated as (Baumann et al. 1997)

$$l_p = l_{p0} + \frac{l_B}{4\kappa^2} = l_{p0} + 3.43I^{-1} \text{nm}, \quad (3.5)$$

where $l_{p0}$ was the non-electrostatic contribution and was set to be 50 nm, $l_B$ was Bjerrum length, $\kappa = 1/\lambda_D$, and $I$ was salt concentration. As shown in figure 8, the persistence length resulting from our computations decreases as a function of the salt concentration. This tendency agrees with experimental analyses and
Figure 8. Persistence length of coarse-grained models as a function of salt concentrations. Agreements between this study and other results explain the validity of potential functions in the present model. Filled circles, present study; open circles, inextensible WLC; open squares, strong stretching limit; open triangles, extensible WLC; dashed lines, Poisson–Boltzmann theory.

the Poisson–Boltzmann theoretical model. However, differences between each model appear at the high salt concentration near 100 mM. Therefore, it can be said that our coarse-graining model explains the behaviour of DNA at low salt concentrations. Reasonable quantities have been obtained from our potential functions, which depend on the charge distribution and elasticity of a DNA fragment. The screening effect represented by the Debye–Hückel approximation is considered to be appropriate. However, the convergence at high salt concentrations appeared to be slower than the experimental and the theoretical estimations. This is because the rotational relaxation times and the distance of base-stack ⟨rise⟩ moderately change as the salt concentration becomes higher. The improvement of the difference with other models is a subject of future study. In the present DNA model, instead of a conventional potential for dihedral angles along the backbones, the stacking potential $V_{\text{stack}}$ that represents the π-stack effect between base pairs was adopted effectively to maintain its double-stranded structure. In base-pair stacks, π–π interactions turn to be effective as electrostatic repulsive interactions between the particles are reduced because of the screening effect. Therefore, differences between the potential functions clearly influence the persistence length that depends on the rotational relaxation time and the distance of base stack ⟨rise⟩. In the present study, it is confirmed that the double-stranded structure is maintained by π–π interactions between the base stacks.

4. Conclusion

In the present study, a coarse-grained model of B–DNA has been developed. The potential function for π-stacks of base pairs and single-nucleotide resolution are particularly emphasized. Using the coarse-grained model, several characteristics have been concluded as follows: (i) the robustness of the double-stranded structure was confirmed in a wide range of salt concentrations compared with experimental
Development of coarse-graining DNA models

(Olson et al. 1998) and other computational (Šponer & Lankaš 2006) results, (ii) it was clarified that the melting properties were dominated by the strength of hydrogen bonds between complementary bases and the melting temperatures agreed well with several DNA sequences, and (iii) the bending properties of DNA were confirmed to be in agreement with the experimental results by investigating the relation between the persistence length and salt concentrations.

References


