Sequence effects in the melting and renaturation of short DNA oligonucleotides: structure and mechanistic pathways

2009 J. Phys.: Condens. Matter 21 034105

(http://iopscience.iop.org/0953-8984/21/3/034105)

J. Phys.: Condens. Matter **21** (2009) 034105 (13pp) doi:10.1088/0953-8984/21/3/034105

Sequence effects in the melting and renaturation of short DNA oligonucleotides: structure and mechanistic pathways

E J Sambriski, V Ortiz and J J de Pablo

Department of Chemical and Biological Engineering, University of Wisconsin-Madison, Madison, WI 53706, USA

E-mail: depablo@engr.wisc.edu

Received 4 July 2008, in final form 28 August 2008
Published 17 December 2008
Online at stacks.iop.org/JPhysCM/21/034105

Abstract
The renaturation/denaturation of DNA oligonucleotides is characterized in the context of expanded ensemble (EXE) and transition path sampling (TPS) simulations. Free energy profiles have been determined from EXE for DNA sequences of varying composition, chain length, and ionic strength. TPS simulations within a Langevin dynamics formalism have been carried out to obtain further information of the transition state for renaturation. Simulation results reveal that free energy profiles are strikingly similar for the various DNA sequences considered in this work. Taking intact double-stranded DNA to have an extent of reaction $\xi = 1.0$, the maximum of the free energy profile appears at $\xi \approx 0.15$, corresponding to ∼2 base pairs. In terms of chain length, the free energy barrier of longer oligonucleotides (30 versus 15 base pairs) is higher and slightly narrower, due to increased sharpness associated with the transition. Low ionic strength tends to decrease free energy barriers, whereby increasing strand rigidity facilitates reassociation. Two mechanisms for DNA reassociation emerge from our analysis of the transition state ensemble. Repetitive sequences tend to reassociate through a non-specific pathway involving molecular slithering. In contrast, random sequences associate through a more restrictive pathway involving the formation of specific contacts, which then leads to overall molecular zippering. In both random and repetitive sequences, the distribution of contacts suggests that nucleation is favored for sites located within the middle region of the chain. The prevalent extent of reaction for the transition state is $\xi \approx 0.25$, and the critical size of the nucleus as obtained from our analysis involves ∼4 base pairs.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

The primary step in the self-assembly of deoxyribonucleic acid (DNA) structures is the association and stabilization of complementary strands. When two complementary nucleic bases (i.e., adenine to thymine and cytosine to guanine) encounter one another with proper orientation, hydrogen bonding leads to base pair association. The double-helix arrangement of DNA arises from base-pairing and base-stacking interactions occurring in tandem along the length of the two complementary strands. The system can be driven away from the associated state by applying heat or by altering solvent conditions, thereby destabilizing the network of nucleic base interactions. In the case of dissociation induced by thermal effects, the system can be brought back to an associated state (renatured or rehybridized) by annealing.

A thermodynamic description of DNA association is of considerable interest. Theory has provided useful frameworks to interpret macroscopic observations of the denaturation cycle in DNA. In constructing analytically tractable formalisms, theoretical models have resorted to presumed mechanisms (particularly the two-state formalism) [1], leaving to question the validity of several underlying assumptions. As some of these formalisms are based on statistical mechanical models of DNA (such as the quasi-one-dimensional Ising model) [2, 3], their ability to provide a molecular understanding of renaturation events is not generally within the scope of such treatments. While reasonable agreement with experimental data can be attained with such models, molecular simulations of more detailed constructs can offer valuable insights pertaining to details that are not included in available analytical treatments.

The renaturation of DNA involves a complex interplay between long- and short-range interactions. A competition between electrostatic repulsion (from negatively charged phosphate sites along the backbone of single-stranded DNA) and medium-induced interactions (such as reorganization of ionic species and solvent molecules) is involved in molecular reassociation. Experimental data on DNA reassociation has been interpreted mostly as a second-order reaction, in which a nucleation step is followed by zippering of the helix [4–7], but the rate-limiting step in this process has been contested [8]. Other aspects of the reassociation, such as the optimal temperature for annealing [9, 10], effect of ionic strength [11], solvent viscosity [12, 13], and pH [14] have also been investigated.

Several theories have been proposed to interpret DNA kinetic experiments [15–19]. Some of these treatments, however, bring to question the assumptions used to characterize the bimolecular reaction. A full understanding of the initiation and size of nucleation in DNA reassociation in several systems is still lacking. For instance, experiments have elucidated essential features of nucleation in the renaturation of large DNA molecules [20, 21]; analogous studies for short oligonucleotides have not been presented. Such information would be relevant for amplification of DNA sequences [22], design of microarray platforms [23], and single-molecule experiments requiring strand control [24]. Molecular simulations could not only help elucidate the overall reassociation reaction, but could also address a number of issues that are fundamental to molecular self-assembly, including those pertaining to oligonucleotides and their nucleation events.

The study of renaturation with atomistic simulations is hindered by two related issues: (a) the disparity in local (atomic) timescales and global (molecular) dynamics, both of which play a role in the process of association, and by (b) the number of degrees of freedom needed to account for electrostatics and solvent effects at such level of detail. Coarse grain models could be used to overcome such issues. Their implementation, however, has been limited because in many cases available constructs cannot describe interactions at the level of individual nucleotides. A description at the level of base pairs, nevertheless, is of interest since important intermediate configurations can be identified in the context of nucleation events needed to describe the mechanism behind molecular association. As a viable alternative to conventional simulation approaches, we developed early on an implicit-solvent, coarse grain model of DNA parameterized with respect to experimental data for mechanical properties and denaturation of short oligonucleotides [25]. The model reduces the number of degrees of freedom in the system, and is able to capture realistically a number of essential physicochemical features of DNA. Most importantly, it resolves interactions at the level of individual nucleotides.

Using an improved version of this model of DNA, in this work we use molecular simulations to study nucleotide sequence effects pertinent to denaturation and renaturation. In particular, we have focused on the characterization of free energy profiles and relevant configurational features leading to the single-to-double-stranded transition. We begin with a brief overview of the DNA model employed here. Next, we describe our implementation of expanded ensembles to study sequence effects in the dis- and reassociation of short oligonucleotides. That study is complemented with transition path sampling simulations aimed at a detailed characterization of the transition state for renaturation.

2. A coarse grain model for DNA

The original DNA construct considered in this work was developed by Knotts *et al* [25]. An improved version of the model was presented by Sambriski *et al* [26]. Such a construct maps each of the three chemical moieties (sugar, phosphate, and nucleic base) of a nucleotide onto three interaction sites. Each of these sites interacts with other sites through a type-specific potential energy function. Standard coordinates obtained from crystallographic data for the B-form, double-stranded DNA (dsDNA) have been used to define a reference molecular structure that governs positional and orientational constraints in the force field. Such an arrangement allows one to capture the biologically relevant feature of major and minor grooves in dsDNA. By adopting an implicit-solvent approach, the model permits study of long DNA chains over long timescales.

The force field for the coarse grain representation of DNA includes intra- and interstrand interactions. The stability of the backbone is partly sustained by intrastrand interactions that control bonds, bends, and torsions. Base-stacking interactions, which are also of an intrastrand nature, are modeled through a Gō-like approach [27]. In this latter contribution, sites interact through a Lennard-Jones potential that depends on pair-specific separations dictated by the reference structure of DNA. All intrastrand sites found within a predefined interaction radius (9 Å) are classified as Gō-like contacts. The effect of hydrogen bonding is incorporated through a short-ranged, Lennard-Jones interaction that acts between complementary nucleic base sites. Base pair mismatches, as well as all other excluded-volume contributions, are treated through a Weeks–Chandler–Anderson potential. Electrostatic contributions are taken into account through phosphate sites, which interact at the level of Debye–Hückel theory. Explicit mathematical expressions for the model, as well as a listing of the corresponding parameters, are provided in the appendix and in the literature [25, 26].

Several limitations in the original formulation of the coarse grain model for DNA [25] have been addressed in the revised model [26]. The original parameter set captured the nominal stiffness of dsDNA (which has an empirical value of 50 nm) to within approximately 50%. The new parameter set for intramolecular interactions (reported in the appendix) remedies this shortcoming by reproducing the nominal value to within an error of approximately 20%. Both former and current forms of the force field capture the dependence of molecular persistence length with ionic strength. The other limitation concerns the denaturation of DNA. Given proper system conditions, complementary strands of single-stranded DNA (ssDNA) can renature to yield dsDNA. While the former representation can describe DNA melting in the context of bubble formation dynamics, there is no mechanism in the force field to capture renaturation within an implicit-solvent construct. The only interstrand attractive contribution in the former force field is hydrogen bonding, which is sharp and short-ranged.

To address the issue of renaturation, the force field now includes a weak, intermediate-range interaction of $\mathcal{O}(10^{-1})\,k_{\mathrm{B}}T$ meant to embody solvent-mediated effects. This contribution takes the form of an interstrand interaction that acts between all sugar site pairs. This follows closely the phenomenology investigated in the context of hydration forces in nucleic acids as shown in experiments by Rau and Persegian [28, 29], the ion-mediated interactions measured by x-ray scattering methods of Qiu and coworkers [30], as well as many-body correlation effects of ionic species discussed in theoretical work by Ha and Liu [31, 32]. The solvent-mediated interaction has been tempered against the attractive contribution of hydrogen bonds to bring about a sufficiently weak effect that can allow for both renaturation and denaturation. We have performed the parameterization of the improved coarse grain representation for DNA using a set of short strands with $n \in \{10, 15, 30\}$, where n denotes the number of nucleotides per single strand. Such a study was performed using a series of oligonucleotides with increasing cytosine–guanine content (f_{CG}) and ionic strength, the results of which are given in a different publication [26].

A Langevin dynamics (LD) formalism is used to generate DNA trajectories. The friction coefficient used for these calculations is derived from diffusivity data for short DNA oligonucleotides. The LD integrator provides for a threefold speedup over the original Nosé–Hoover implementation. Details on these aspects of the model can be found elsewhere [26].

3. Free energy via expanded ensembles

The method of expanded ensembles (EXE) facilitates the study of thermodynamic properties of a system as a function of a suitably defined order parameter [33–36]. Provided the order parameter can distinguish between relevant phases of the system, expanded ensembles are particularly effective for studies of phase transitions. In this work, the number of bound base pairs is selected as the order parameter to characterize the denaturation transition in DNA. At constant N, V, and T, an expanded ensemble is defined through the relation

$$\Omega = \sum_{i=0}^{n} Q(N, V, T, \xi_i)\mathrm{e}^{\Upsilon_i}, \tag{3.1}$$

where $Q(N, V, T, \xi_i)$ denotes the canonical ensemble partition function corresponding to a particular state ξ_i of the total $(n + 1)$ states of the system. As described below, one can design an algorithm that samples successive states ξ_i of the system sequentially. In that case, ξ_i can be used to define the 'extent of reaction' of a particular process. For each ξ_i, there exists an exponentiated positive weighting factor Υ_i such that each $Q(N, V, T, \xi_i)$ contributes equally to Ω [37]. The probability of finding the system in state ξ_i is given by

$$p(\xi_i) = \frac{Q(N, V, T, \xi_i)}{\Omega}\mathrm{e}^{\Upsilon_i}. \tag{3.2}$$

From equation (3.2), it follows that

$$\frac{\tilde{p}(\xi_i)}{\tilde{p}(\xi_n)} = \frac{Q(N, V, T, \xi_i)}{Q(N, V, T, \xi_n)}, \tag{3.3}$$

with $\tilde{p}(\xi_i) \equiv p(\xi_i)\mathrm{e}^{-\Upsilon_i}$. The optimal weighting factors Υ_i are such that all ξ_i states are sampled with equal probability; these optimal weights acquire an important physical significance, namely, their differences are associated with changes in the Helmholtz free energy A of the system. Expanded ensemble formalisms therefore allow for a systematic calculation of the free energy as a function of the extent of reaction $A(\xi)$. Using equation (3.2), the corresponding difference in free energy can be conveniently defined as a ratio of partition functions with respect to some reference state ξ_0 as

$$\begin{aligned}\Delta A_{0\to i} &\equiv A(\xi_i) - A(\xi_0)\\ &= -k_{\mathrm{B}}T \ln\left[\frac{Q(N, V, T, \xi_i)}{Q(N, V, T, \xi_0)}\right] = -k_{\mathrm{B}}T \ln\left[\frac{\tilde{p}(\xi_i)}{\tilde{p}(\xi_0)}\right].\end{aligned} \tag{3.4}$$

Values for $p(\xi_i)$ can be obtained by evolving the expanded ensemble with any method of choice, provided it is adept at permitting the system to traverse configurational space efficiently. At the end of a simulation cycle, each $p(\xi_i)$ is extracted from a histogram of visits to each state ξ_i.

Trial perturbations or moves are performed on the molecule to sample each of the $(n + 1)$ states. When a trial move is proposed, it is accepted or rejected with probability

$$P_{\mathrm{acc}}(\xi_i \to \xi_j) = \min\left[1, \frac{T(\xi_j \to \xi_i)}{T(\xi_i \to \xi_j)}\frac{p(\xi_j)}{p(\xi_i)}\right], \tag{3.5}$$

where $T(\xi_i \to \xi_j)$ is the probability of proposing a transition from state ξ_i to ξ_j, and $p(\xi_i)$ is the probability of observing the system in state ξ_i. Transition probabilities in equation (3.5) must account for any asymmetry in the system. In traversing a one-dimensional order parameter space, end states are asymmetrically disposed to transition between neighboring states when compared to intermediate ones, and the ratio of transition probabilities is in general not unity [35]. Whenever a proposed move violates an imposed boundary (relative or absolute) for a chosen interval of states, the move is rejected by updating the histogram of visited states of the originating state [35, 36, 38].

With the aid of equation (3.3), the canonical ensemble $p(\xi_i)$ can be evaluated, so that for any two configurations ξ_i and ξ_j of the system,

$$\frac{p(\xi_j)}{p(\xi_i)} = \frac{\mathrm{e}^{-\beta U_j}\mathrm{e}^{\Upsilon_j}}{\mathrm{e}^{-\beta U_i}\mathrm{e}^{\Upsilon_i}}, \tag{3.6}$$

where $\beta = (k_BT)^{-1}$. Using equation (3.6), equation (3.5) simplifies to

$$P_{\text{acc}}(\xi_i \to \xi_j) = \min\left[1, \mathrm{e}^{-\beta(U_j-U_i)}\mathrm{e}^{(\Upsilon_j-\Upsilon_i)}\right]. \tag{3.7}$$

In effect, equation (3.7) not only accounts for the change in energy as the system evolves from ξ_i to ξ_j, but also for any change in the order parameter of the expanded ensemble, thereby facilitating uniform sampling of all states. In this sense, the evolution of the system corresponds to a one-dimensional Markovian trajectory of the ξ_i states.

In the context of DNA denaturation, our EXE simulation is envisioned as a series of $(n + 1)$ states pooled from $\xi_i \in \{\xi_0 \cdots \xi_n\}$, with n representing the maximum number of native interstrand base pairs that can form for a given sample of dsDNA. The intact state of dsDNA is denoted by ξ_n, while ξ_0 represents the fully molten state involving ssDNA. All ξ_i such that $\xi_0 < \xi_i < \xi_n$ correspond to states with an intermediate extent of reaction. Our EXE calculation aims to bridge smoothly between the ξ_0 and ξ_n states. To investigate the renaturation/denaturation cycle, we define our EXE thermodynamic path as that in which base pairs form (or break) sequentially from one end of the molecule to the other (i.e., only native contacts are allowed). This arrangement builds the zippering mechanism commonly alluded to in the literature into our simulation algorithm [25]. To prepare a system for an EXE simulation, we allow for an equal number of configurations to be partially denatured from the 3′- and the 5′-end.

To expedite convergence, we partition the range of $(n + 1)$ states into a set of sub-representations (or 'windows') consisting of several consecutive values of ξ. Each window encompasses a small range of ξ, or a small number of states. Each window is populated with a configuration of two complementary DNA strands acquired from an equilibrated canonical MD simulation. All windows are evolved simultaneously. Each window is allowed to overlap with several states from adjacent windows; attempts to swap configurations from adjacent windows are performed at a preset frequency. An exchange occurs if states from two windows overlap when a swap move is proposed. The system is evolved using a hybrid Monte Carlo–molecular dynamics (MCMD) method [39]. This approach samples configurational space with randomly alternating intervals of Monte Carlo and molecular dynamics simulations. The Monte Carlo scheme employed here includes translation, rotation, and pivot trial moves [40]. The MCMD method facilitates sampling of configurational space in both a local and global manner.

The weighting factors Υ_i are acquired in two stages from a series of simulation runs that progressively increase in length. During the first stage of sampling ('priming'), a short simulation run is performed with all weighting factors set to a common value. The number of steps in this simulation is sufficiently large to allow the system to sample several states in each window (for our case, $\mathcal{O}(10^5)$ steps). When complete, a probability is computed from the histogram of visits to each state. These probabilities serve to determine the Υ_i through the relation

$$\Upsilon_i^{\text{new}} = \Upsilon_i^{\text{old}} - \ln[p(\xi_i)]. \tag{3.8}$$

Initially, all weighting factors are set to zero ($\Upsilon_i^{\text{old}} = 0$) in equation (3.8). States with no hits [$p(\xi_i) = 0$] are assigned an interpolated weight involving neighboring states. In the second stage of sampling ('refinement'), the Υ_i^{new} serve as input to longer runs (for our studies, these refinement runs are of $\mathcal{O}(10^6)$ steps). This process is repeated until probabilities satisfy a desired statistical accuracy (for our case, when all $p(\xi_i)$ deviate less than 15% from their nominal value). The difference in excess free energy is determined from equation (3.4). Before equation (3.4) can be used, information from all the windows that sample the entire range of ξ space must be combined. The optimal additive constant of each window for Υ_i can be determined through the weighted histogram analysis method (WHAM) [41, 42],

$$\varrho(\xi_i) = C\frac{\sum_{\alpha=1}^{N_w} h_\alpha(\xi_i)}{\sum_{\gamma=1}^{N_w} H_\gamma \mathrm{e}^{-[\Upsilon_\gamma(\xi_i)-F_\gamma]}}, \tag{3.9}$$

$$\mathrm{e}^{-F_\gamma} = C\sum_{i=0}^{n} \varrho(\xi_i)\mathrm{e}^{-\Upsilon_\gamma(\xi_i)}, \tag{3.10}$$

where N_w is the number of windows used in the EXE simulation and $\varrho(\xi_i)$ is a discretized probability distribution. For the αth simulation, $h_\alpha(\xi_i)$ denotes the number of counts accrued by state ξ_i, H_α is the total number of data points collected, and F_α is an optimized additive constant. The F_α are found by iterating equations (3.9) and (3.10) until self-consistency is achieved. The convergence of an EXE calculation can be assessed by the extent to which weights from overlapping regions agree once the F_α are applied.

Information from one thermodynamic state point can be extrapolated to a nearby state point by extending equation (3.2). This is particularly useful when seeking the precise conditions under which a phase transition occurs. More specifically, if information is available at a reference state point β_0, and additional information at point β is required, the extrapolation takes the form [43, 44]:

$$\tilde{p}_\beta(\xi_i) = \frac{\tilde{p}_{\beta_0}(\xi_i)\sum_{g=1}^{N_b}\mathrm{e}^{-(\beta-\beta_0)U_g(\xi_i)}}{\sum_{i=0}^{n}\tilde{p}_{\beta_0}(\xi_i)\sum_{h=1}^{N_b}\mathrm{e}^{-(\beta-\beta_0)U_h(\xi_i)}}, \tag{3.11}$$

where $\tilde{p}(\xi_i)$ is defined by equation (3.3), while N_b is the number of bins used to discretize the potential energy $U(\xi_i)$. In the case of two-state systems, the change in free energy is expected to vanish at a phase transition since the probability for either state is identical,

$$\Delta A_{0\to m} \equiv A(\xi_m) - A(\xi_0) = -k_BT\ln\left[\frac{\tilde{p}(\xi_m)}{\tilde{p}(\xi_0)}\right] = 0, \tag{3.12}$$

where ξ_m is another state with a free energy minimum. By combining the extrapolation of equation (3.11) with the condition of equation (3.12), one can locate phase transitions with considerable precision.

4. Transition path sampling and the transition state ensemble

In the study of phase transitions, much information can be gleaned about a process based on the characterization of intermediate states or extents of reaction. The principal step in the reassociation of ssDNA involves a nucleation event. While information on the nucleation during the renaturation of large DNA systems has been provided experimentally [20, 21], the manner in which it occurs and the critical size of the nucleus for smaller oligonucleotides has not been considered. Such information would lead to a better understanding of the process of DNA association, and would enable the design of protocols aimed at presenting molecules in specific, optimal arrangements for renaturation or rehybridization.

Such questions can be addressed through methods aimed at harvesting information about a reaction on the basis of transitions, which is the viewpoint adopted in transition path sampling (TPS). TPS requires that a 'primitive path' be identified. That path is a trajectory (in the case of MD simulations, coordinates and momenta) in which the system goes from one state to another. From this trajectory, one adds perturbations at distinct time slices to generate a collection of reactive trajectories (i.e., those which successfully traverse a path going from one established state to another). Such trajectories constitute the transition path ensemble (TPE), which can then be analyzed to obtain thermodynamic properties of the system involving transitions between states of interest.

As this work focuses on the problem of the single-to-double-stranded transition in DNA, it is convenient to establish a convention to denote the intact or fully denatured states. As with our EXE calculations, our order parameter is given by the extent of reaction ξ. To characterize the system in terms of a chemical reaction, we define the *forward reaction* to be DNA renaturing, and a range of states is defined to be in region B when $\xi \rightarrow 1.0$. The *reverse reaction* involves the breaking of hydrogen bonds between complementary base pairs to yield ssDNA, and a range of states is defined to be in region A when $\xi \rightarrow 0.0$.

To sample the space of transition paths and collect the TPE, we use shooting moves in which one selects a time slice at random from the 'primitive path' trajectory (in our case, for a trajectory of S steps, our initial time slice t_0 with configuration X_0 is selected from the trajectory fragment $0.3S \leqslant t_0 \leqslant 0.7S$). To this time slice, a perturbation is added, which for our case amounts to assigning new momenta. A 'forward trajectory' ($X_0 \rightarrow B$) is then grown from $t_0 \leqslant t \leqslant S$. If the system fails to situate itself in region B during the time interval $(S - t_0)$, the proposed time slice is rejected. On the other hand, if the system arrives at region B, a backward leg is constructed by inverting the momenta at t_0 and a 'backward trajectory' ($X_0 \rightarrow A$) is generated from $0 \leqslant t < t_0$ in an attempt to reach region A. If this backward trajectory successfully reaches region A, the entire trajectory is accepted, otherwise the time slice is rejected and another time slice is chosen at random.

Once the TPE has been collected, we proceed to generate the transition state ensemble (TSE). This ensemble is comprised by the set of configurations belonging to a surface in configurational space where trajectories are equally likely to end in either region A or B [45]. In keeping with our original definition of a chemical reaction, our collection of trajectories is selected by determining the extent of reaction ξ (i.e., fraction of base pair contacts) that a given configuration acquires. To identify these configurations from the TPE, one selects points along a given trajectory and performs short, fleeting trajectories. These trajectories are then used to determine committor probabilities [45], defined by the expression,

$$\pi_{\mathrm{B}} = \frac{1}{N_f} \sum_{i=1}^{N_f} h_{\mathrm{B}}^i. \tag{4.1}$$

Here, N_f denotes the total number of fleeting trajectories, while h_{B}^i represents a counting function for region B of the ith trajectory. If a fleeting trajectory successfully connects a state in region A (molten DNA) with one in region B (fully renatured DNA), then $h_{\mathrm{B}}^i = 1$, otherwise $h_{\mathrm{B}}^i = 0$. To categorize whether a reaction has gone to completion, we have defined a threshold of $\xi = 0.8$ for the extent of reaction. All configurations that start from some intermediate state of bound DNA and are consistently characterized by $\xi = 0.8$ for at least one-quarter of a fleeting trajectory have $h_{\mathrm{B}}^i = 1$. For our studies, a fleeting trajectory spans a timescale of 2 ns out of the 20 ns trajectories from the TPE.

By using equation (4.1), the TSE is defined by the set of configurations for which $\pi_{\mathrm{B}} = 0.5$. We define a tolerance with respect to π_{B} through the relation [46]

$$\sigma = \left[\frac{\pi_{\mathrm{B}}(1 - \pi_{\mathrm{B}})}{N_f}\right]^{1/2}, \tag{4.2}$$

where N_f represents the minimum number of fleeting trajectories needed for statistical significance. For our work, we have used $N_f = 100$ and with the condition that $\pi_{\mathrm{B}} = 0.5$, the minimum tolerance yields $\sigma = 0.05$. A configuration is designated as being part of the TSE if $[\pi_{\mathrm{B}} - \nu\sigma, \pi_{\mathrm{B}} + \nu\sigma]$, where ν is chosen to satisfy a desired confidence interval. We have used a confidence interval of 95% ($\nu = 2$), such that a configuration is part of the TSE whenever $0.4 < \pi_{\mathrm{B}} < 0.6$.

5. Results

5.1. Systems of study

The oligonucleotides considered in this study include chain lengths of $n \in \{14, 15, 30\}$ base pairs, with several chain compositions f_{CG}, as summarized in table 1. The conditions investigated here are given in table 2. The EXE melting temperature simulations have been performed for two chain lengths $n \in \{15, 30\}$ at all compositions, while for TPS, only the $n \in \{14, 15\}$ cases have been investigated, each with an associated f_{CG}. Taken together, these systems enable us to address a number of effects relevant to chain length, ionic strength, and nucleotide sequence. All MD simulations were performed with a Langevin dynamics integrator in the canonical ensemble.

5.2. Expanded ensemble

The melting of DNA is perceived to be a first-order phase transition [49–51]. The melting temperature, T_{m}, is often identified as the point at which $\xi = 0.5$, as

Table 1. Sequences of short oligonucleotide series.

n	f_{CG}	Sequence (5′–3′)									
14	0.0	ATA	TAT	ATA	TAT	AT					
15	0.0	ATA	TAT	ATA	TAT	ATA					
15	0.2	TAC	TAA	CAT	TAA	CTA					
15	0.4	AGT	AGT	AAT	CAC	ACC					
30	0.0	ATA	TAT	ATA	TAT	ATA	TAT	ATA	TAT	ATA	TAT
30	0.2	TTA	TGT	ATT	AAG	TTA	TAT	AGT	AGT	AGT	AGT

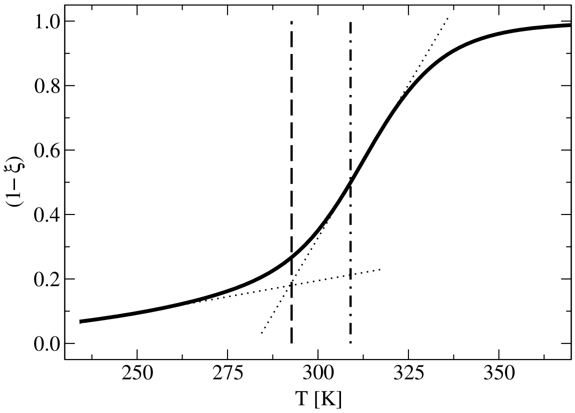

Figure 1. Comparison of denaturation temperatures from a melting curve (solid line) where the fraction of melting base pairs $(1 - \xi)$ is plotted as a function of temperature T. Shown are data for $n = 15$ ($f_{CG} = 0.2$), which has an empirical melting temperature of $T_m = 308.4$ K. The corresponding results from EXE (dashed line) yields $T_m = 296$ K while REMD (dot-dashed line) yields at the half-point $T_m = 309$ K. The alternate definition for T_m is given by the slopes depicted (dotted lines), whose intersection defines the melting temperature at the onset of denaturation.

Table 2. T_m for short oligonucleotide series.

n	f_{CG}	[Na$^+$] M	$T_{m,obs}$	$T_{m,REMD}$[a]	$T_{m,EXE}$
14	0.0	0.069	291.2[b]	—	—
15	0.0	0.069	294.8[b]	—	282
15	0.2	0.005	283.6[b]	—	275
15	0.2	0.069	308.4[c]	295	296
15	0.4	0.069	317.4[c]	308	310
30	0.0	0.069	312.9[b]	—	308
30	0.2	0.069	323.8[c]	316	318

[a] T_m for the onset of denaturation. The half-point T_m is reported in [26]. Also see figure 1.

[b] Data from [48].

[c] Data from [47].

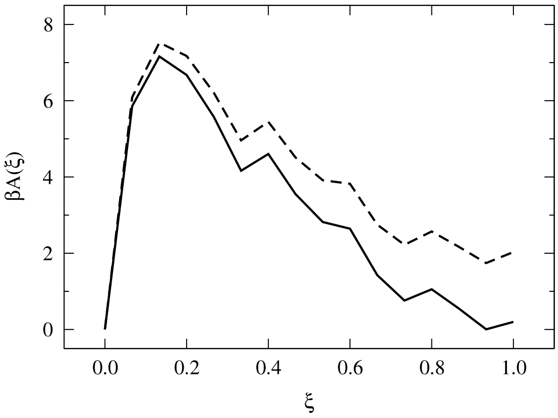

Figure 2. Determination of the thermodynamic melting temperature. Shown is the free energy profile as a function of ξ, for the system in figure 1, using the half-point melting temperature from REMD (dashed curve) and that determined by setting $\Delta A = 0$ (solid curve), consistent with the alternate definition for T_m depicted in figure 1.

determined with a melting curve obtained from UV absorbance measurements [1]. A thermodynamic definition, however, would assign T_m to the state point in which the difference between two minima of free energy vanishes. From this perspective, it is of interest to compare these definitions for T_m. Our studies are also prompted by experimental melting profiles of short oligonucleotides [52, 47, 53], where the melting transition is relatively broad, and an ambiguity arises in the definition of T_m.

In order to establish a correspondence between different techniques for studying melting, we compared the melting transition temperature as determined from previous studies using a replica exchange molecular dynamics (REMD) algorithm [26] with that obtained from EXE simulations. While we only present results for $n = 15$ ($f_{CG} = 0.2$), similar findings hold for other systems. Using $\xi = 0.5$ to define T_m, the melting temperature can be determined by inspection of the denaturation curve obtained from REMD. Such a temperature can be used as an input for EXE calculations. The EXE simulations are then carried out until two 'minima' display equal heights, at which point the system is characterized by $\Delta A = 0$. Figure 1 shows a comparison between the melting temperature obtained from EXE and REMD simulations. A discrepancy of about 10 K is evident: the EXE calculation exhibits a melting temperature of 296 K, while the temperature from REMD at the half-point is 309 K. The former was found by using equation (3.11) and solving by iteration for the state point at which the minima of the free energy profile are of equal weight, the results of which are shown in figure 2. This state point was confirmed with an EXE simulation performed at the same conditions (not shown). The origin of this difference can be attributed to the broadness of the transition associated for a relatively small system. Indeed, the ambiguity of defining the melting temperature vanishes for larger systems, where the transition as given by REMD displays a sharper response [26]. EXE simulations indicate that for short oligonucleotides, a thermodynamically consistent way of defining T_m from melting curves is to identify it with the onset of melting, as indicated by the intersection of the slopes (dotted lines) shown in figure 1.

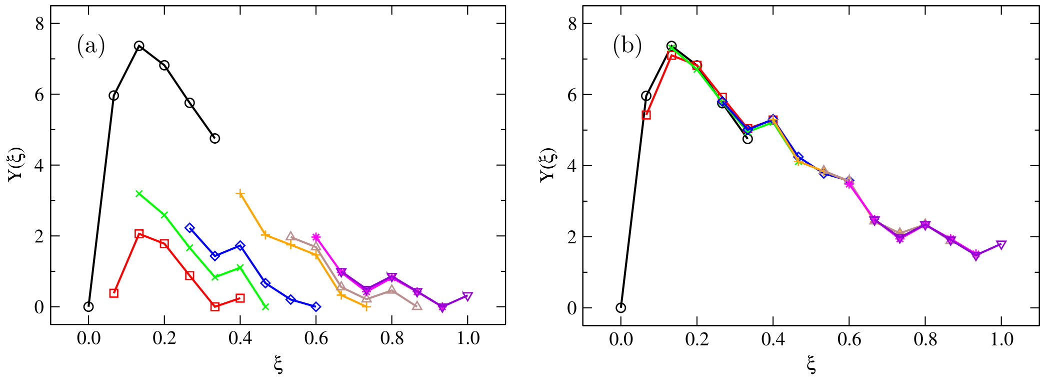

Figure 3. Implementation of WHAM to combine results for the weighting factors $\Upsilon(\xi)$ from multiple simulation windows in the determination of free energy profiles. The raw data (panel (a)) and the data after applying the additive constants (panel (b)) are shown.

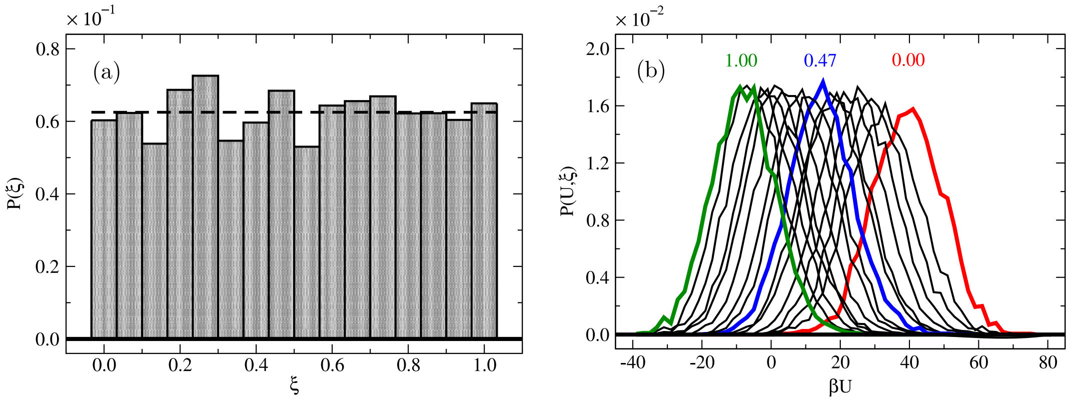

Figure 4. Verification of uniform sampling. Data are shown for the normalized histogram of visits (panel (a)) and potential energy distributions (panel (b)) for each state. For panel (a), the corresponding ideal probability is also shown (dashed line). For panel (b), ξ decreases from left to right; some values of ξ for $P(U, \xi)$ are highlighted (heavy, colored lines).

Simulating the melting transition and renaturation of DNA poses a number of sampling challenges. A few details concerning the validity of EXE simulations are therefore in order. Figure 3(a) addresses the extent of overlap among weights from neighboring windows. After applying equations (3.9), (3.10) and shifting the weights by their respective additive constants, the profile of weights yields the results shown in figure 3(b). These findings help establish that the sampling between windows is statistically sound, and that the error associated for a given $\Upsilon(\xi_i)$ is of $\mathcal{O}[10^{-1}]$. Uniform sampling can also be gauged from the corresponding energy profiles and histograms of visits to each state ξ_i. These results are shown in figure 4, where each state is visited with equal probability within our predefined tolerance (panel 4(a)) and characterized by an energy distribution comparable to that of neighboring states (panel 4(b)).

Figure 5 shows free energy profiles of all sequences for $n = 15$. While each system displays a unique signature (albeit weak) in its free energy function, a striking finding from the simulations is the qualitative similarity of the distinct profiles at the melting transition. Results for the $n = 30$ case are shown in figure 6. For these systems, the overall shape of the free energy function is qualitatively similar, but the values of the free energy are higher and the profile is slightly narrower than that observed for the shorter molecules, attesting to the increased sharpness of the associated transition. Figure 7 depicts the free energy and entropy landscapes for a spectrum of temperatures as a function of DNA base pair contacts for $f_{\mathrm{CG}} = 0.0$ (panels 7(a) and (b)) and $f_{\mathrm{CG}} = 0.4$ (panels 7(c) and (d)). As can be gleaned from these figures,

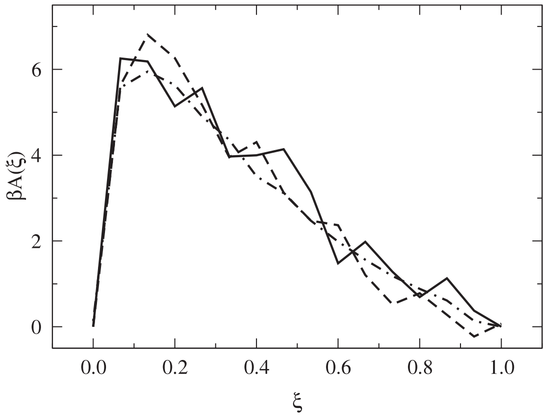

Figure 5. Free energy profiles for different systems. Data are shown for $n = 15$ with $f_{\mathrm{CG}} = 0.0$ (dot–dashed line), $f_{\mathrm{CG}} = 0.2$ (dashed line), and $f_{\mathrm{CG}} = 0.4$ (solid line).

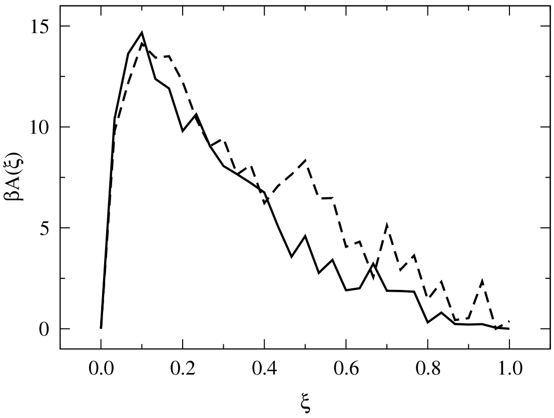

Figure 6. Same as in figure 5, but for $n = 30$ with $f_{CG} = 0.0$ (dashed line) and $f_{CG} = 0.2$ (solid line).

the process of renaturation/denaturation exhibits a significant dependence on entropic contributions. In particular, at higher temperatures entropic contributions become substantial. States corresponding to low ξ exhibit invariably higher entropic contributions, since in this limit the constraints of hydrogen bonding are inhibited and molecules can explore a large space of disassociated configurations.

The model considered in this work has been shown to be capable of describing ionic effects on melting temperature and on persistence length. It is therefore of interest to consider the effect of salt on the transition of the system for $n = 15$ ($f_{CG} = 0.2$). Note that so far, all EXE calculations were performed at $[\mathrm{Na}^+] = 0.069$ M, as shown by the solid line in figure 8. The corresponding free energy profile for $[\mathrm{Na}^+] = 0.005$ M is shown by the dashed line, which exhibits $T_m = 275$ K. At this salt concentration, energy barriers for molecular reassociation are lowered by increasing molecular stiffness. Figure 8 suggests that more 'extended' or open configurations allow two complementary strands to present themselves in a manner that facilitates renaturation. Molecular strengthening has been suggested as the mechanism behind the phenol emulsion reassociation technique (PERT), in which phenol stabilizes ssDNA in an extended, rod-like conformation [54]. While the context in which renaturation occurs in PERT is different from the present study, the parallel to draw is that molecular stretching of ssDNA favors reassociation.

We end this section with a few remarks concerning the actual process of renaturation, as inferred from the free energy profiles obtained from EXE simulations. To gain some preliminary insights into the nature of the critical nucleus size and its role in system dynamics, we focus on the $n = 15$ ($f_{CG} = 0.2$) system and examine the appearance of the free energy in figure 8. From that figure, the critical nucleus size is found to be $n\xi = 2$. Note, however, that our EXE simulations were conceived to follow a specific (and in some sense arbitrary) pathway. That pathway does not necessarily correspond to the actual pathway followed by complementary strands when they renature. We therefore investigated the preference that states located near this region of the free energy profile have in sampling either basin (which we assign to states $n\xi = 0$ and 14, respectively). We performed fleeting LD trajectories (for a timescale of 200 ps) on configurations from prior EXE simulations characterized by $n\xi \in \{1, 2, 3\}$. For

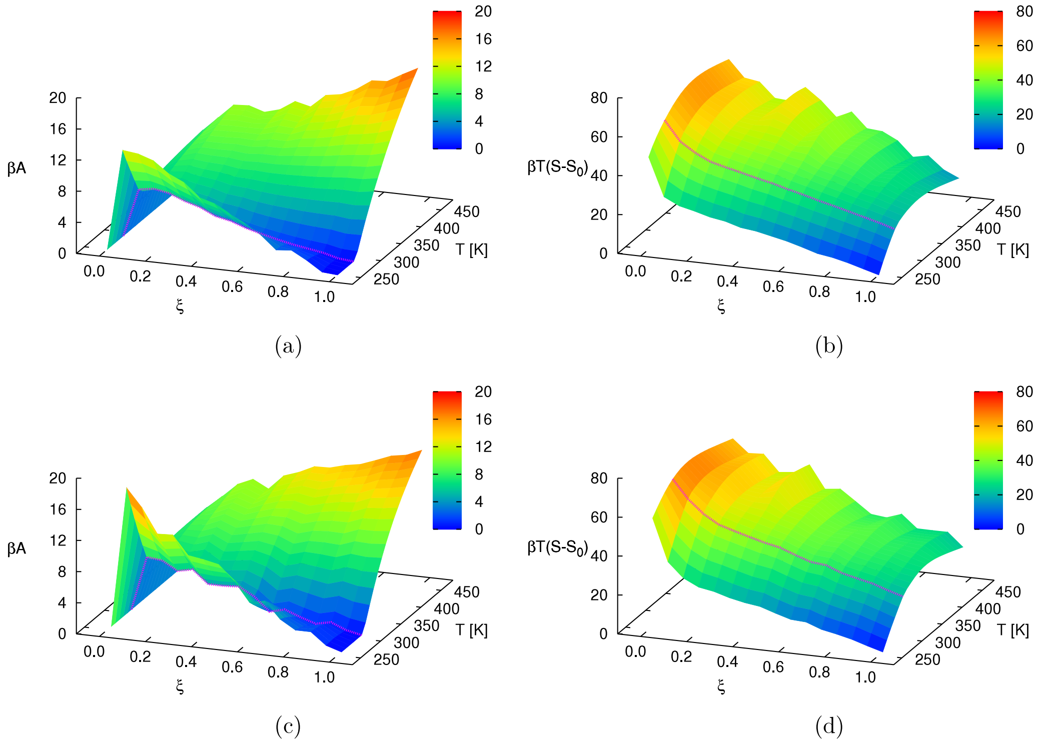

Figure 7. Free energy (βA) landscapes (panels (a) and (c)) and entropy $[\beta T(S - S_0)]$ landscapes (panels (b) and (d)) as a function of ξ and T. Data are shown for $n = 15$ with $f_{CG} = 0.0$ (panels (a) and (b)) and $f_{CG} = 0.4$ (panels (c) and (d)). The profile at the melting temperature for each system is denoted with a magenta line. S_0 corresponds to the entropy of the system for intact dsDNA ($\xi = 1.0$) for the lowest temperature shown.

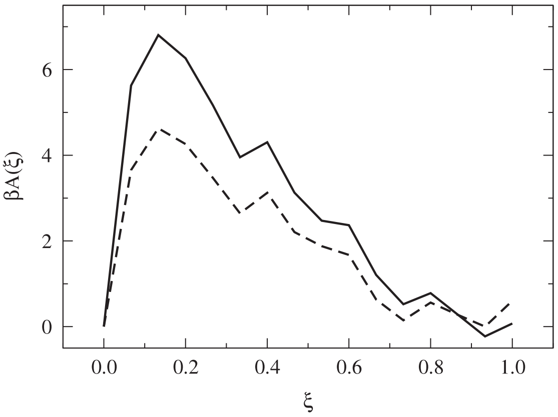

Figure 8. Salt effects on the free energy profile. Results are shown for $n = 15$ ($f_{\mathrm{CG}} = 0.2$) with $[\mathrm{Na}^+] = 0.005$ M (dashed line) and 0.069 M (solid line).

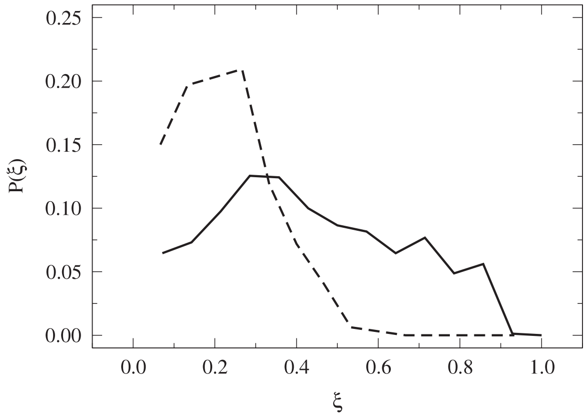

Figure 9. Probability of extents of reaction in the TSE. Results are shown for $f_{\mathrm{CG}} = 0.0$ (solid line) and $f_{\mathrm{CG}} = 0.2$ (dashed line).

Table 3. Characterization of the critical nucleus.

Extent of reaction, ξ	Fraction configurations with $\xi \to 1$
0.06	0.42
0.13	0.59
0.20	0.69

each value of ξ_i, ten initial configurations were considered (half-denatured from the 5′-end, the other half from the 3′-end), each of which was subjected to 100 fleeting trajectories. From these trajectories, runs which display a traversal tending to $\xi \to 1.0$ were counted, the results of which are displayed in table 3. We find that roughly half of the configurations with $n\xi = 2$ display a route leading to the fully renatured state. These results suggest that $n\xi = 2$ could correspond to a critical nucleus size for renaturation. Within the assumptions or constraints that went into the EXE thermodynamic cycle, the size of the nucleus already establishes a preference for the system to follow an association pathway. The association mechanism inferred from those LD trajectories tending to $\xi = 1.0$ exhibits sequential reassociation, consistent with the thermodynamic pathway followed by our EXE simulations.

5.3. Transition path sampling

As mentioned above, the predefined thermodynamic pathway assumed for our EXE calculations suggests the existence of a critical nucleus for renaturation of two base pairs. That thermodynamic pathway, however, was imposed on the calculations to extract a free energy. It is therefore of interest to investigate and compare those results with those extracted from TPS simulations, in which no renaturation mechanism is assumed *a priori*. TPS calculations were performed at ~5 K below $T_{\mathrm{m,obs}}$ (see table 2) for $n = 15$ ($f_{\mathrm{CG}} = 0.2$), and $n = 14$ ($f_{\mathrm{CG}} = 0.0$) (the equivalent $n = 15$ system is less interesting from symmetry considerations). We refer to these systems as ‘random’ and ‘repetitive’ sequences, respectively. We analyzed in detail configurations belonging to the TSE (see section 4 for further details on the TSE definition). We determined the probability to find a specific extent of reaction $P(\xi)$, the results of which are shown in figure 9. The distribution for the $f_{\mathrm{CG}} = 0.0$ case is rather broad, involving a ‘non-specific’ pathway, while that of the $f_{\mathrm{CG}} = 0.2$ case clearly displays distinct signatures for association.

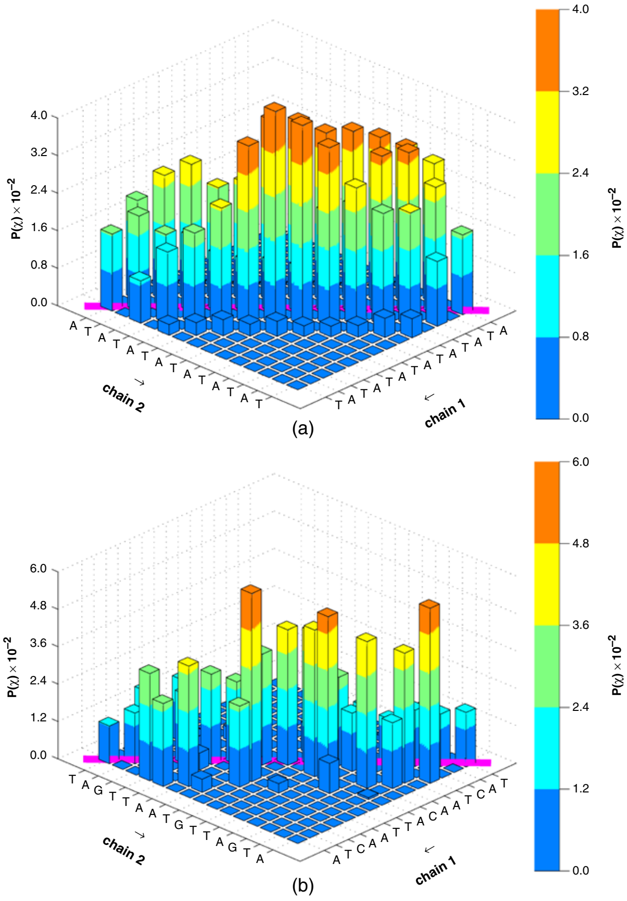

Figure 10. Joint probability of base pair contacts $P(\chi)$ for all extents of reaction ξ shown in figure 9. Data are shown for $f_{\mathrm{CG}} = 0.0$ (panel (a)) and $f_{\mathrm{CG}} = 0.2$ (panel (b)). The sequence for each strand is denoted along the axes, with ‘chain 1’ being the sense strand. Arrows along the axes denote the 5′-to-3′ direction for each chain. Native base pair contacts are denoted along the diagonal (magenta) line.

To investigate further the nature of the transition state for the reassociation of DNA, we determined from the TSE the

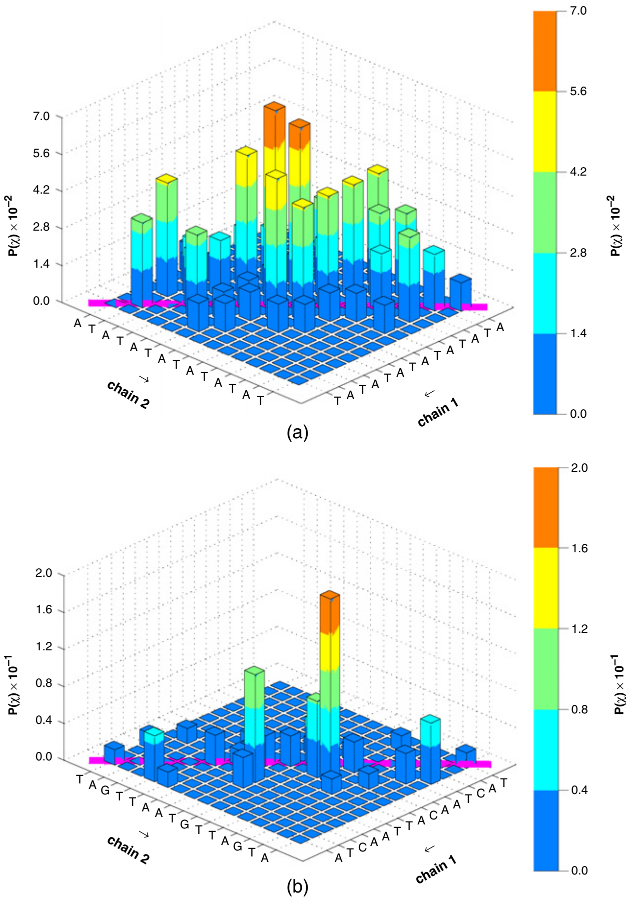

Figure 11. The same as in figure 10, but for $n\xi = 1$.

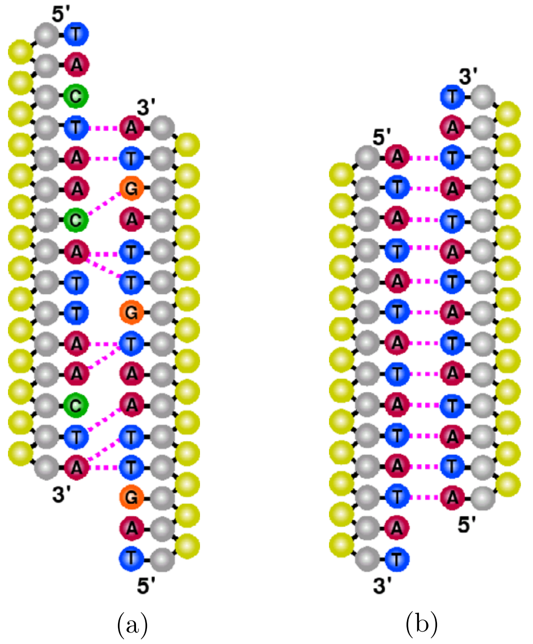
5'
T A C T A A C A T T A A C T A
3'
3'
A T G A T T G T A A T T G A T
5'
(a)
5'
A T A T A T A T A T A T A T
3'
3'
T A T A T A T A T A T A T A
5'
(b)

Figure 12. Schematic of DNA systems identifying base pairs (dotted, magenta lines) with highest probability in the TSE. Shown are characteristic reassociation sites for $f_{\mathrm{CG}} = 0.2$ (panel (a)) and $f_{\mathrm{CG}} = 0.0$ (panel (b)). The sense strand for each case is placed on the left. Moieties are denoted by color as sugar (gray), phosphate (yellow), adenine (red), cytosine (green), guanine (orange), and thymine (blue) sites.

joint probability that complementary base pairs will associate, $P(\chi)$. Those distributions are shown in figure 10. Native contacts in that figure correspond to diagonal entries. For the $f_{\mathrm{CG}} = 0.2$ case, one observes that a large portion of TSE configurations display the formation of a few key contacts. On the other hand, the same distribution for $f_{\mathrm{CG}} = 0.0$ is less specific. Further decomposition of this distribution into the formation of one base pair is given in figure 11. It is evident that for early extents of reaction, key base pair contacts happen primarily in the middle portion of the chain. As the system nears $\xi \sim 0.5$, an increasingly large fraction of contacts form near-native base pairs in both the random and the repetitive sequences.

Figure 12 provides a schematic representation of configurations that arise with a high contact probability in the TSE. These figures do not necessarily account for the sequence in which base pairs form, but rather help elucidate preferred regions for molecular association. In the case of the random sequence, only specific contacts are highly prevalent. This suggests that for random DNA, molecular reassociation occurs via a selective, sequence-dependent pathway. For the repetitive sequence, on the other hand, the reassociation pathway is less selective: many contacts exhibit similar probabilities. Figure 11 indicates that the repetitive system still exhibits a preference for initial nucleation involving central base pairs. The progression of figure 11 from one to four bonds (not shown) and the connectivity of base pairs (not shown) reveals that molecular slithering plays a stronger role in repetitive DNA than in the random sequence.

6. Summary

A study on the denaturation/renaturation cycle of DNA has been performed in the context of expanded ensemble (EXE) and transition path sampling (TPS) simulations. Free energy profiles acquired from EXE for systems prepared with identical conditions but varying composition exhibit qualitatively similar profiles. For short oligonucleotides, it is found that the thermodynamic melting transition corresponds to the onset of denaturation curves, as opposed to the mid range of such curves. The maximum in the free energy occurs for small extents of reaction, $\xi \approx 0.15$, corresponding to ~2 base pairs. The profile for larger chains displays a higher and slightly narrower response, indicative of the sharper nature of the transition. As salt concentration is decreased, a lowering of the free energy profile is observed. This effect presumably arises because molecular reassociation is facilitated by an increase in molecular rigidity.

At the melting temperature, an analysis of the transition state, as acquired by TPS, reveals two mechanisms for molecular reassociation. Initial contacts tend to involve base pairs near the middle region of the sequence, yielding a transition state in which complementary strands are offset from one another by two to three bases. In the case of repetitive sequences, reassociation occurs through a non-specific pathway in which a nucleation event is followed by molecular slithering. For random sequences, the pathway is more sequence specific and involves the formation of key contacts, which subsequently facilitate molecular zippering. At the melting temperature, the critical nucleus size obtained from analysis of TPS data is $\xi \approx 0.25$, corresponding to ~4 base pairs. Future work will seek to characterize the critical nucleus for chains of different lengths, a broader range of compositions, and for different degrees of supercooling.

Acknowledgments

EJS was supported by an NHGRI training grant to the Genomic Sciences Training Program (5T32HG002760). VO was supported by an NLM training grant to the Computational and Informatics in Biology and Medicine Training Program (NLM T15LM007359). This work was also supported by the National Science Foundation through the University of Wisconsin-Madison NSEC.

Appendix. Coarse grain force field for DNA

The force field for the coarse grain model of DNA consists of both bonded and non-bonded interactions. These contributions are described below with relevant numerical parameters listed in table A.1. The total contribution of the potential energy of the system is given by the sum of eight distinct contributions,

$$U_{\text{tot}} = U_{\text{bond}} + U_{\text{bend}} + U_{\text{tors}} + U_{\text{stck}} + U_{\text{base}} + U_{\text{nnat}} + U_{\text{elec}} + U_{\text{solv}}. \tag{A.1}$$

Contributions from 'bonded' interactions are given by two-, three-, and four-body terms, which represent bonding, bending, and torsion constraints, respectively, through the expressions

$$U_{\text{bond}} = \sum_{i=1}^{n_{\text{bond}}} \left[k_1 (d_i - d_{0i})^2 + k_2 (d_i - d_{0i})^4 \right], \tag{A.2}$$

$$U_{\text{bend}} = \sum_{i=1}^{n_{\text{bend}}} \frac{k_\theta}{2} (\theta_i - \theta_{0i})^2 , \tag{A.3}$$

$$U_{\text{tors}} = \sum_{i=1}^{n_{\text{tors}}} k_\phi \left[1 - \cos(\phi_i - \phi_{0i})\right]. \tag{A.4}$$

Here, U_{bond} accounts for covalent bonding, with k_1 and k_2 representing bond constants, while d_i and d_{0i} are instantaneous and equilibrium site–site separations for the ith bond in the set of n_{bond} bonds. Molecular bending is accounted for through U_{bend}, with a bend constant k_θ, as well as an instantaneous and equilibrium bond angle θ_i and θ_{0i} for the ith bend in the set of n_{bend} bends. Torsional interactions are represented through U_{tors}, with a torsional constant k_ϕ, along with instantaneous and equilibrium dihedral angles ϕ_i and ϕ_{0i} in the set of n_{tors} torsions. To preserve the right-handed chirality in B-form DNA, the scheme of Hoang and Cieplak [27] is used.

Non-bonded, pairwise interactions are described through five contributions as

$$U_{\text{stck}} = \sum_{i<j}^{n_{\text{stck}}} 4\varepsilon \left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6} \right], \tag{A.5}$$

Table A.1. Force field parameters.

Parameter	Value	Units
ε	0.769 856	kJ mol^{-1}
ε_{AT}	2.000ε	kJ mol^{-1}
ε_{CG}	2.532ε	kJ mol^{-1}
ε_{s}	System-dependent	kJ mol^{-1}
k_1	ε	kJ mol^{-1} Å^{-2}
k_2	100ε	kJ mol^{-1} Å^{-4}
k_θ	1400ε	kJ mol^{-1} rad^{-2}
k_ϕ	28ε	kJ mol^{-1}
α^{-1}	5.333	Å
r_{s}	13.38	Å
σ_{ij}	Pair-dependent	Å
σ_{AT}	2.9002	Å
σ_{CG}	2.8694	Å
σ_0 (mismatch)	$1.00 \times 2^{-1/6}$	Å
σ_0 (otherwise)	$6.86 \times 2^{-1/6}$	Å

$$U_{\text{base}} = \sum_{i=1}^{n_{\text{base}}} 4\,\varepsilon_{\text{b}i} \left[5\left(\frac{\sigma_{\text{b}i}}{r_{ij}}\right)^{12} - 6\left(\frac{\sigma_{\text{b}i}}{r_{ij}}\right)^{10} \right], \tag{A.6}$$

$$U_{\text{nnat}} = \sum_{i<j}^{n_{\text{nnat}}} \begin{cases} 4\,\varepsilon \left[\left(\frac{\sigma_0}{r_{ij}}\right)^{12} - \left(\frac{\sigma_0}{r_{ij}}\right)^{6} \right] + \varepsilon & \text{if } r_{ij} < r_{\text{coff}} \\ 0 & \text{if } r_{ij} \geqslant r_{\text{coff}}, \end{cases} \tag{A.7}$$

$$U_{\text{elec}} = \sum_{i<j}^{n_{\text{elec}}} \frac{q_i q_j \mathrm{e}^{-r_{ij}/\lambda_{\text{D}}}}{4\pi\epsilon_0 \epsilon(T, I) r_{ij}}, \tag{A.8}$$

$$U_{\text{solv}} = \sum_{i<j}^{n_{\text{solv}}} \varepsilon_{\text{s}} \left[1 - \mathrm{e}^{-\alpha(r_{ij} - r_{\text{s}})}\right]^2 - \varepsilon_{\text{s}}. \tag{A.9}$$

The first two contributions are specific to DNA systems. Base stacking (an intrastrand effect) is taken into account through U_{stck}, which acts uniformly (i.e., with a single energy scale ε) on all native contacts n_{stck} (i.e., intrastrand sites found within a cutoff radius $r_{\text{ccut}} = 9$ Å in the reference structure of DNA). An interaction specific length scale σ_{ij} between sites i and j contributes to the stiffness of the DNA backbone by controlling the instantaneous site–site separation r_{ij}.

Hydrogen bonding is accounted for by U_{base}, and acts between all n_{base} complementary (Watson–Crick) base pairs that do not participate in U_{stck} (i.e., this term takes into account both intra- and interstrand possibilities). Each ith base pair, characterized by the separation r_{ij} between intra- or interstrand sites i and j, will be governed by energies $\varepsilon_i \in \{\varepsilon_{\text{AT}}, \varepsilon_{\text{CG}}\}$ and lengths $\sigma_i \in \{\sigma_{\text{AT}}, \sigma_{\text{CG}}\}$, where $\varepsilon_{\alpha\beta} = \varepsilon_{\beta\alpha}$ and $\sigma_{\alpha\beta} = \sigma_{\beta\alpha}$. A complementary base pair is considered to be hydrogen bonded when the separation between bases is $r_{ij} < (\sigma_i + 2.0$ Å).

Interactions occurring between the n_{nnat} non-native contacts (including mismatched base pairs) are assigned to U_{nnat}. This is a purely repulsive, excluded-volume contribution (a Weeks–Chandler–Anderson interaction) characterized by a single energy scale ε. An energy cost arises in the system as interparticle separations r_{ij} fall below a cutoff length scale r_{coff}, but otherwise U_{nnat} does not contribute to the stability of the system. In the case of mismatched base pairs, $r_{\text{coff}} = 1.00$ Å, while in all other cases, $r_{\text{coff}} = 6.86$ Å, the latter of which corresponds to a mean pair separation value. The length scale at which the potential vanishes is given by $\sigma_0 = 2^{-1/6} r_{\mathrm{coff}}$.

The remaining two non-bonded interactions supply the additional physics needed to model 'polyelectrolyte' features of DNA. Electrostatic contributions are treated through U_{elec} at the level of Debye–Hückel theory. This term accounts for interactions involving all possible n_{elec} pairings between phosphate sites i and j that do not enter into U_{bend}. The Debye length, λ_{D}, which defines the spatial extent of charge screening at an interparticle separation r_{ij} due to solvent conditions, is given by

$$\lambda_{\mathrm{D}} = \left[\frac{\epsilon_0 \epsilon(T, I)}{2\beta N_{\mathrm{A}} e_{\mathrm{c}}^2 I}\right]^{1/2}, \qquad (\mathrm{A.10})$$

where ϵ_0 is the permittivity of free space, $\epsilon(T, I)$ is an effective dielectric constant, $\beta = (k_{\mathrm{B}} T)^{-1}$, N_{A} is Avogadro's number, e_{c} is the elementary charge, and I is the ionic strength of the solution (identical to the molarity of the solution for the 1:1 electrolyte NaCl considered in this study). The dielectric constant $\epsilon(T, I)$, specialized here to treat aqueous salt solutions, includes a dependence on temperature and salt concentration, both of which have a direct bearing on the polarizability of water. Decomposition of the effective dielectric constant into a product of contributions was suggested earlier in the literature [55] as

$$\epsilon(T, I) = \epsilon(T) a(I), \qquad (\mathrm{A.11})$$

where $\epsilon(T)$ is the static (zero-frequency) dielectric constant at absolute temperature T (in kelvin), and $a(I)$ is the salt correction for a solution with molarity I in NaCl. Each of these contributions, in turn, are given by

$$\epsilon(T) = 249.4 - 0.788\, T/\mathrm{K} + 7.20 \times 10^{-4}\, (T/\mathrm{K})^2, \qquad (\mathrm{A.12})$$

and

$$a(I) = 1.000 - 0.2551\, I/\mathrm{M} + 5.151 \times 10^{-2}\, (I/\mathrm{M})^2 - 6.889 \times 10^{-3}\, (I/\mathrm{M})^3. \qquad (\mathrm{A.13})$$

The solvent-induced contribution, U_{solv}, is a novel addition to the force field meant to embody (implicitly) the entropic effects associated with the arrangement of water and ionic species during denaturation or renaturation of DNA. This Morse-like interaction is characterized by an energy scale ε_{s} and an interparticle separation r_{ij}, while the spatial range of the potential is controlled by α^{-1}, with a potential energy minimum occurring at r_{s}. Solvent-induced interactions n_{solv} enter into U_{solv} through all possible interstrand sugar site pairings. Parameters for this contribution are chosen to be compatible with the molecular geometry of DNA. As the solvent-induced interaction is meant to embody many-body effects, it depends on chain length and ionic conditions. Parameterization of the model against empirical data yields the following expressions for the strength of interaction as a function of chain length,

$$\varepsilon_{\mathrm{N}} = \varepsilon_{\infty} \left(1 - [1.404\,18 - 0.268\,231 n]^{-1}\right), \qquad (\mathrm{A.14})$$

with $\varepsilon_{\infty} = 0.504\,982\varepsilon$, while the salt dependence is approximated by the expression,

$$A_{\mathrm{I}} = 0.474\,876 \left(1 + \{0.148\,378 + 10.9553[\mathrm{Na}^+]\}^{-1}\right). \qquad (\mathrm{A.15})$$

The two effects from equations (A.14) and (A.15) can be combined provided $n \gtrsim 15$ through the relation $\varepsilon_{\mathrm{s}} \approx A_{\mathrm{I}} \varepsilon_{\mathrm{N}}$, with A_{I} serving as a universal curve.

References

[1] Wartell R M and Benight A S 1985 Thermal denaturation of DNA molecules: a comparison of theory with experiment *Phys. Rep.* **126** 67–107

[2] Hill T L 1959 Generalization of the one-dimensional Ising model applicable to helix transitions in nucleic acids and proteins *J. Chem. Phys.* **30** 383–7

[3] Lifson S and Zimm B H 1963 Simplified theory of the helix–coil transition in DNA based on a grand partition function *Biopolymers* **1** 15–23

[4] Marmur J, Rownd R and Schildkraut C L 1963 Denaturation and renaturation of deoxyribonucleic acid *Prog. Nucleic Acid Res.* **1** 231–300

[5] Rau D C and Klotz L C 1978 A unified theory of nucleation-rate-limited DNA renaturation kinetics *Biophys. Chem.* **8** 41–51

[6] Murugan R 2002 Sample size autocorrelation analysis of kinetic data: resolving reaction path heterogeneity *J. Chem. Phys.* **117** 6372–7

[7] Britten R J and Davidson R H 1976 Studies on nucleic acid reassociation kinetics: empirical equations describing DNA reassociation *Proc. Natl Acad. Sci. USA* **73** 415–9

[8] Rau D C and Klotz L C 1975 A more complete kinetic theory of DNA renaturation *J. Chem. Phys.* **62** 2354–65

[9] Marmur J and Doty P 1961 Thermal renaturation of deoxyribonucleic acids *J. Mol. Biol.* **3** 585–94

[10] Nygaard A P and Hall B D 1964 Formation and properties of RNA–DNA complexes *J. Mol. Biol.* **9** 125–42

[11] Marmur J and Lane D 1960 Strand separation and specific recombination in deoxyribonucleic acids: biological studies *Proc. Natl Acad. Sci. USA* **46** 453–61

[12] Subirana J A and Doty P 1966 Kinetics of renaturation of denaturated DNA. I. Spectrophotometric results *Biopolymers* **4** 171–87

[13] Thrower K J and Peacocke A R 1966 The kinetics of renaturation of DNA *Biochim. Biophys. Acta* **119** 652–4

[14] Wetmur J G 1976 Hybridization and renaturation kinetics of nucleic acids *Annu. Rev. Biophys. Bioeng.* **5** 337–61

[15] Saunders M and Ross P D 1960 A simple model of the reaction between polyadenylic acid and polyuridylic acid *Biochem. Biophys. Res. Commun.* **3** 314–8

[16] Kallenback N R, Crothers D M and Mortimer R G 1963 Interpretation of the kinetics of helix formation *Biochem. Biophys. Res. Commun.* **11** 213–6

[17] Murugan R 2003 A stochastic model on DNA renaturation kinetics *Biophys. Chem.* **104** 535–41

[18] Murugan R 2003 A theory on the origin of cooperativity in DNA renaturation kinetics *Biophys. Chem.* **106** 173–8

[19] Murugan R 2002 Revised theory on DNA renaturation kinetics and its experimental verification *Biochem. Biophys. Res. Commun.* **293** 870–3

[20] Matsuzawa Y, Yonezawa Y and Yoshikawa K 1996 Formation of nucleation center in single double-stranded DNA chain *Biochem. Biophys. Res. Commun.* **225** 796–800

[21] Yoshikawa K and Matsuzawa Y 1996 Nucleation and growth in single DNA molecules *J. Am. Chem. Soc.* **118** 929–30

[22] Saiki R K, Gelfand D H, Stoffel S, Scharf S J, Higuchi R, Horn G T, Mullis K B and Erlich H A 1988 Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase *Science* **239** 487–91

[23] Buck M J and Lieb J D 2004 ChIP–chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments *Genomics* **83** 349–60

[24] Ramanathan A, Huff E J, Lamers C C, Potamousis K D, Forrest D K and Schwartz D C 2004 An integrative approach for the optical sequencing of single DNA molecules *Anal. Biochem.* **330** 227–41

[25] Knotts T A IV, Rathore N, Schwartz D C and de Pablo J J 2007 A coarse grain model for DNA *J. Chem. Phys.* **126** 084901

[26] Sambriski E J, Schwartz D C and de Pablo J J 2009 Mesoscale model of DNA and its renaturation *Biophys. J.* at press

[27] Hoang T X and Cieplak M 2000 Molecular dynamics of folding of secondary structures in Gō-type models of proteins *J. Chem. Phys.* **112** 6851–62

[28] Rau D C and Parsegian V A 1992 Direct measurement of the intermolecular forces between counterion-condensed DNA double helices: evidence for long range attractive hydration forces *Biophys. J.* **61** 246–59

[29] Rau D C and Persegian V A 1992 Direct measurement of temperature-dependent solvation forces between DNA double helices *Biophys. J.* **61** 260–71

[30] Kwok L W, Qiu X, Park H Y, Lamb J S, Andresen K and Pollack L 2006 Measuring inter-DNA potentials in solution *Phys. Rev. Lett.* **96** 138101

[31] Ha B-Y and Liu A J 1998 Charge oscillations and many-body effects in bundles of like-charged rods *Phys. Rev.* E **58** 6281–6

[32] Ha B-Y and Liu A J 1999 Counterion-mediated, non-pairwise-additive attractions in bundles of like-charged rods *Phys. Rev.* E **60** 803–13

[33] Lyubartsev A P, Martsinovski A A, Shevkunov S V and Vorontsov-Velyaminov P N 1992 New approach to Monte Carlo calculation of the free energy: method of expanded ensembles *J. Chem. Phys.* **96** 1776–83

[34] Wilding N B and Müller M 1994 Accurate measurements of the chemical potential of polymeric systems by Monte Carlo *J. Chem. Phys.* **101** 4324–30

[35] Escobedo F A and de Pablo J J 1995 Monte Carlo simulation of the chemical potential of polymers in an expanded ensemble *J. Chem. Phys.* **103** 2703–10

[36] Escobedo F A and de Pablo J J 1996 Expanded grand canonical and Gibbs ensemble Monte Carlo simulation of polymers *J. Chem. Phys.* **105** 4391–4

[37] de Pablo J J, Yan Q and Escobedo F A 1999 Simulation of phase transitions in fluids *Annu. Rev. Phys. Chem.* **50** 377–411

[38] Shulz B J, Binder K, Müller M and Landau D P 2003 Avoiding boundary effects in Wang–Landau sampling *Phys. Rev.* E **67** 067102

[39] Rathore N, Knotts T A IV and de Pablo J J 2003 Density of states simulations of proteins *J. Chem. Phys.* **118** 4285–90

[40] Rathore N and de Pablo J J 2002 Monte Carlo simulation of proteins through a random walk in energy space *J. Chem. Phys.* **116** 7225–30

[41] Roux B 1995 The calculation of the potential of mean force using computer simulations *Comput. Phys. Commun.* **91** 275–82

[42] Souaille M and Roux B 2001 Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations *Comput. Phys. Commun.* **135** 40–57

[43] Ferrenberg A M and Swendsen R H 1988 New Monte Carlo technique for studying phase transitions *Phys. Rev. Lett.* **61** 2635–8

[44] Chopra M, Müller M and de Pablo J J 2006 Order-parameter-based Monte Carlo simulation of crystallization *J. Chem. Phys.* **124** 134102

[45] Bolhuis P G, Chandler D, Dellago C and Geissler P L 2002 Transition path sampling: throwing ropes over rough mountain passes, in the dark *Annu. Rev. Phys. Chem.* **53** 291–318

[46] Dellago C, Bolhuis P G and Geissler P L 2002 Transition path sampling *Adv. Chem. Phys.* **123** 1–78

[47] Owczarzy R, You Y, Moreira B G, Manthey J A, Huang L, Behlke M A and Walder J A 2004 Effects of sodium ions on DNA duplex oligomers: improved predictions of melting temperatures *Biochemistry* **43** 3537–54

[48] http://www.idtdna.com/analyzer/Applications/OligoAnalyzer

[49] Garel T, Monthus C and Orland H 2001 A simple model for DNA denaturation *Europhys. Lett.* **55** 132–8

[50] Carlon E, Orlandini E and Stella A L 2002 Role of stiffness and excluded volume in DNA denaturation *Phys. Rev. Lett.* **88** 198101

[51] Kafri Y, Mukamel D and Peliti L 2000 Why is the DNA denaturation transition first-order? *Phys. Rev. Lett.* **85** 4988–91

[52] Holbrook J A, Capp M W, Saecker R M and Record M T Jr 1999 Enthalpy and heat capacity changes for formation of an oligomeric DNA duplex: interpretation in terms of coupled processes of formation and association of single-stranded helices *Biochemistry* **38** 8409

[53] Owczarzy R, Vallone P M, Gallo F J, Paner T M, Lane M J and Benight A S 1997 Predicting sequence-dependent melting stability of short duplex DNA oligomers *Biopolymers* **44** 217–39

[54] Goldar A and Sikorav J-L 2004 DNA renaturation at the water–phenol interface *Eur. Phys. J.* E **14** 211–39

[55] Stogryn A 1971 Equations for calculating the dielectric constant of saline water *IEEE Trans. Microw. Theory Tech.* **19** 733–6