1872–1885 Nucleic Acids Research, 2021, Vol. 49, No. 4 *Published online 27 January 2021*
doi: 10.1093/nar/gkab015

Dinucleotides as simple models of the base stacking-unstacking component of DNA 'breathing' mechanisms

Eric R. Beyerle[1,†], Mohammadhasan Dinpajooh[1,†], Huiying Ji[2,3], Peter H. von Hippel[3], Andrew H. Marcus[2,3] and Marina G. Guenza[1,2,3,*]

[1]Institute for Fundamental Science and Department of Chemistry and Biochemistry, University of Oregon, Eugene, OR 97403, USA, [2]Center for Optical, Molecular and Quantum Science and Department of Chemistry and Biochemistry, University of Oregon, Eugene, OR 97403, USA and [3]Institute of Molecular Biology and Department of Chemistry and Biochemistry, University of Oregon, Eugene, OR 97403, USA

Received October 22, 2020; Revised December 22, 2020; Editorial Decision January 04, 2021; Accepted January 07, 2021

ABSTRACT

Regulatory protein access to the DNA duplex 'interior' depends on local DNA 'breathing' fluctuations, and the most fundamental of these are thermally-driven base stacking-unstacking interactions. The smallest DNA unit that can undergo such transitions is the dinucleotide, whose structural and dynamic properties are dominated by stacking, while the ion condensation, cooperative stacking and inter-base hydrogen-bonding present in duplex DNA are not involved. We use dApdA to study stacking-unstacking at the dinucleotide level because the fluctuations observed are likely to resemble those of larger DNA molecules, but in the absence of constraints introduced by cooperativity are likely to be more pronounced, and thus more accessible to measurement. We study these fluctuations with a combination of Molecular Dynamics simulations on the microsecond timescale and Markov State Model analyses, and validate our results by calculations of circular dichroism (CD) spectra, with results that agree well with the experimental spectra. Our analyses show that the CD spectrum of dApdA is defined by two distinct chiral conformations that correspond, respectively, to a Watson–Crick form and a hybrid form with one base in a Hoogsteen configuration. We find also that ionic structure and water orientation around dApdA play important roles in controlling its breathing fluctuations.

INTRODUCTION

Nucleic acids undergo a variety of local structural fluctuations in discharging their biological functions. These fluctuations (collectively called 'breathing') include inter-strand base-pair opening and closing, intra-strand base stacking and unstacking and conformational rearrangements of the sugar-phosphate backbone (1–6). Such thermally activated DNA 'breathing' fluctuations are thought to represent primary steps in the process by which genome regulatory proteins gain access to the double-stranded (ds) DNA interior.

Understanding thermally driven DNA fluctuations may provide a central key to structural and dynamic interpretation of the interactions between functional and regulatory proteins and their ss- and dsDNA targets during gene expression. However, many of these 'breathing' processes, if considered only in duplex DNA, are likely to represent a small fraction of the population of conformations present in duplex DNA at physiological temperatures because of 'structural cooperativity' and may thus be hard to resolve even by sensitive spectroscopic and computational techniques. One way of reducing this problem is to focus on elementary systems, such as dinucleotides. These can be considered to represent the 'fundamental fragments' of duplex DNA, but also provide a milieu in which the only relevant breathing process is likely to be base stacking and unstacking. As a consequence, these processes can be studied in isolation in these small model systems. In addition, because of the absence of constraints imposed by neighboring and base-paired nucleotides, these stacking-unstacking fluctuations are likely to be present at higher concentrations than in larger duplex DNA molecules and thus also more amenable to study. These considerations have motivated us to reinvestigate the structure and dynamics of dApdA as a model

*To whom correspondence should be addressed. Tel: +1 541 3462877; Fax: +1 541 3464643; Email: mguenza@uoregon.edu
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.
Present address: Mohammadhasan Dinpajooh, Department of Chemistry, University of Pennsylvania, Philadelphia, PA 19104, USA.

© The Author(s) 2021. Published by Oxford University Press on behalf of Nucleic Acids Research.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Downloaded from https://academic.oup.com/nar/article/49/4/1872/6121468 by guest on 01 July 2021

dinucleotide fragment of duplex DNA using modern computational and molecular modeling techniques.

The relative populations of stacked and unstacked bases present in DNA molecules in solution under a variety of environmental conditions have traditionally been studied by absorbance and circular dichroism (CD) experiments (7,8). Initial studies of DNA stacking-unstacking fluctuations focused on dinucleotides in solution (9–15). Dinucleotides, such as dApdA, have been considered to be useful models for some of the basic interactions that control and stabilize local base conformations of dsDNA because—as indicated above—stacking interactions can be examined in these systems while avoiding the complicating features of ion condensation, cooperative stacking and inter-base hydrogen-bonding that are also present and involved in controlling the conformational behavior of long duplex DNA. In addition, homo-dinucleotides, such as dApdA, are more useful than hetero-dinucleotides as model systems for probing conformational rearrangements in these structures because the CD signals from homo-dinucleotides are strengthened by the presence of degenerate exciton coupling effects. Furthermore, dinucleotides may also serve as partial models for deciphering the structure and energetics of some of the more complex elements of biologically important DNA structure, such as the single-stranded (ss) DNA–dsDNA forks and junctions that are essential intermediates in the pathways by which proteins that control genome expression find and interact with their target sites, but in which cooperative interactions and hydrogen bonds between strands are not significantly present.

Base and base-pair interaction free energies have typically been estimated from thermal denaturation studies of DNA oligonucleotides (16,17), which showed that among the contributions to the overall interaction free energies of these systems, the free energy of hydrogen bonding between complementary bases and the energetics of configurational and solvent entropy provide only small contributions to the stability of the base paired structures (14). Furthermore, base-base stacking, which is the main (enthalpic) contributor to the stability of dinucleotide conformations, appears also to be the dominant component of the overall stability of more complex DNA structures (10,18,19).

Early studies of dApdA by Schellman *et al.* (8,10,12–14,18) suggested that the CD spectrum of this dinucleotide in aqueous salt solutions could be represented as the weighted sum of two conformations, one 'stacked' and the other 'unstacked', with the stacked form likely resembling (in terms of base-base overlap and helical pitch) the Watson–Crick B-form characteristic of duplex DNA. Furthermore, these workers showed that the changes induced in the CD spectrum of this dinucleotide by increasing concentrations of monovalent salt (NaCl) could be attributed to shifts in the relative populations of these same two conformations.

However, these interpretations clearly represented over-simplifications of the actual situation, since we now know that the CD spectrum of a given molecule of this sort must comprise a sum over myriad microstate configurations that exist simultaneously in solution at equilibrium. As a consequence of this complex situation, CD spectra cannot be 'inverted' to determine the conformations that contribute uniquely to the overall spectrum. We here address this problem by means of extensive Molecular Dynamics (MD) simulations and a Markov State Model (MSM) analysis (20–22), thus providing information on the major conformations that participate in the stacking-unstacking equilibria of dApdA, and whose excitonic transitions contribute to the overall CD spectrum.

To this end, we performed a set of 2 μs MD simulations of the dApdA dinucleotide in aqueous solvent at increasing monovalent salt concentrations, using the same conditions employed for the initial spectroscopic measurements on dApdA dinucleotide (15). From our MD trajectories, each consisting of $\sim 10^7$ microstate configurations, we calculated the CD spectrum by averaging together the contributions from each MD-generated conformation using the standard method (11,12,23,24) together with an extended dipole model (EDM) (24). The initial predictions generated by this method are in good general agreement with experimental spectra previously measured by others.

We next carried out an MSM analysis of our MD trajectories and identified five kinetically stable regions in the free energy landscape, which we refer to as 'macrostates.' Each macrostate contains a 'family' of conformationally-related microstates, which rapidly interconvert. Transitions between macrostates are kinetically uncoupled, because they are separated by high energy barriers and thus follow Markovian statistics (20). The ensemble of macrostates provides a structural basis that can be used to interpret the experimental spectroscopic measurements. By combining MSM analyses with transition path sampling (25–29) we investigated the kinetic pathways for base unstacking, thus revealing the roles that base 'flipping' appears to play in breathing fluctuations at the dinucleotide level. In addition, we were able to identify one average configuration for each macrostate that served, with sufficient accuracy, to represent the averaged properties of the macrostate. This simplified, five-configuration model retains the important features of the CD spectrum calculated using the full MD statistics and provides a useful minimalistic ensemble for the calculation of CD and potentially also other optical spectra obtained using more sophisticated experimental techniques.

Of the five macrostates, three are statistically the most populated, with the CD spectrum being largely determined as the sum of contributions from only two configurational states, consistent with early experimental observations (10). While the original studies interpreted those spectra in terms of a single stacked and a single unstacked configuration of dApdA, our analysis shows that, of the two conformational states that contribute significantly to the features of the CD spectra, the most populated corresponds to an ensemble of hybrid dinucleotide conformations that include one base that has flipped into a *syn* conformation, which in dsDNA results in Hoogsteen base-pairing (30–37), while the relatively less populated state corresponds to an ensemble in which both bases of the dinucleotide remain in the canonical *anti* conformation, compatible with right-handed B form (Watson–Crick) base-pairing in dsDNA. The third highly populated dApdA conformation, which is partially unstacked and contains one *syn* base, does not contribute significantly to the CD signal. However, these results do indicate that conformations compatible with the Hoog-

Downloaded from https://academic.oup.com/nar/article/49/4/1872/6121468 by guest on 01 July 2021

sten structure could well play an important role in some types of breathing fluctuations—at least at the dinucleotide level—thus confirming its possible relevance to biologically important breathing fluctuations in larger DNA molecules as well (30–37).

Our studies of the orientations and distributions of counterions in aqueous solutions of dApdA have revealed an abrupt structural transition in the positioning and distribution of these ions around the dinucleotide at a NaCl concentration slightly above 1 M, indicating that counterion concentrations are also involved in controlling breathing fluctuations at the dinucleotide level (19,38), and thus likely to play a role in the 'breathing' of larger DNA molecules as well. We show that the above abrupt salt-concentration-dependent transition is correlated with a shift in the equilibria between the three most populated macrostates of the dApdA dinucleotide, and is consistent with early thermal studies of DNA stability at increasing monovalent salt concentration (39–41). We have shown that this transition is not seen in MD simulations of the isolated phosphate anion in ionic solutions, suggesting that this salt-dependent transition depends also on other (uncharged) components of the dinucleotide structure.

MATERIALS AND METHODS

Molecular dynamics (MD) simulations

MD simulations of the dApdA dinucleotide monophosphate molecule in aqueous solution were performed at increasing salt concentrations ([NaCl] = 0.1, 0.5, 1.0, 1.05, 1.2 and 1.5 M) in the NPT ensemble using the GROMACS software program (42). The length of the simulation box was allowed to fluctuate, so that the average distance between the box boundary and the dApdA molecule was ∼20 Å. The initial configuration for the dApdA dinucleotide was selected as the B-form conformation, for which we obtained atomic coordinates from the ambertools software package (http://casegroup.rutgers.edu/). Simulations were performed with the Amber03 force-field (43) and the TIP3P water model (44) to model the dApdA molecule and the water component of the solvent, respectively. While these models were not specifically parameterized to achieve accurate CD calculations of the dApdA dinucleotide, they have been used successfully for nucleic acid systems in the past and represent present state-of-the-art for simulations of DNA in solution. A sufficient number of sodium and chloride ions were included to achieve the target salt concentration. The energy of the solvated structure was minimized using the Steepest Descent algorithm for 500 steps. The system was then heated to 300 K and equilibrated in the isothermal–isobaric (NPT) ensemble using a time step of 2 fs over a period between 50 and 100 ns.

Production runs at each salt concentration were performed for a total duration of 2 μs in the NPT ensemble in order to ensure sufficient sampling of the conformational landscape. These simulations used the stochastic velocity-rescaling thermostat (45) with a time constant of 0.2 ps, and the Parrinello-Rahman barostat (using an isotropic pressure coupling time constant of 1.0 ps). We implemented the Leap-Frog algorithm to integrate Newton's equations of motion using the LINCS constraints fourth order in the expansion of the constraint coupling matrix, which included one iteration to correct for rotational lengthening (46). We set the time step to 2 fs, and truncated the Lennard–Jones interactions using a cutoff distance of 10.0 Å. We additionally used a particle mesh Ewald sum to handle long-range electrostatic interactions with a real space cutoff of 10.0 Å and a grid spacing of 1.0 Å. The Verlet neighbor list algorithm was applied with a frequency of 10 MD steps to enhance the computational speed. Trajectory frames were stored every 0.2 ps. In Figure 1, we show a sample frame from one of our MD trajectories. At each salt concentration we included ∼10 million such frames in our CD calculations for the dApdA system.

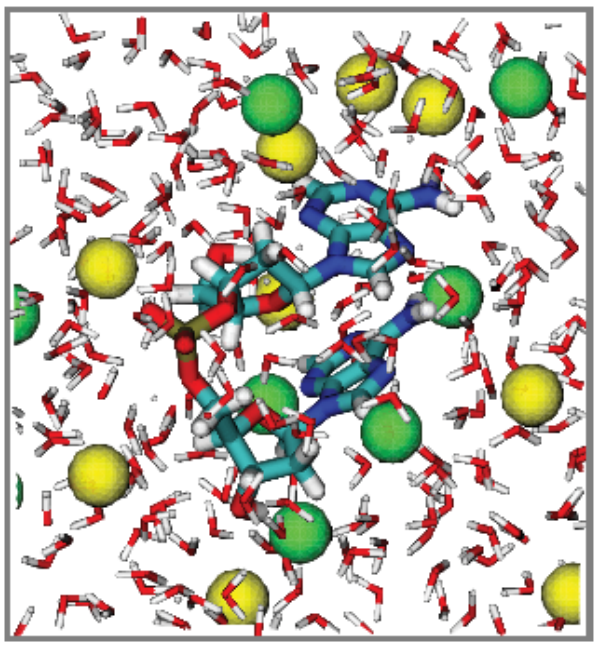

Figure 1. A sample configuration frame taken from an MD simulation run of dApdA dinucleotide monophosphate in TIP3P water with [NaCl] = 0.1 M. Sodium ions are shown as yellow spheres, and chloride ions as green spheres. The atoms of the dApdA and water are colored according to CPK rules, except for carbon, which is colored light blue.

Theoretical modeling of circular dichroism (CD)

The CD spectra for the dApdA dinucleotide monophosphate were calculated from the molecular coordinates of each simulation frame by summing over the contributions from each individual k electronic transition, according to $\Delta\varepsilon(\nu) = \sum_{k=1}^{n_{tot}} \Delta\varepsilon(\nu_k)$. For each of the k electronic transitions, we approximate the CD spectral line shape as a Gaussian function $\Delta\varepsilon(\nu_k) = \Delta\bar{\varepsilon}_k exp\{-[(\nu_k - \bar{\nu}_k)^2/2\sigma_k^2]\}$, where σ_k is the Gaussian standard deviation, $\bar{\nu}_k$ ($= E_k/h$) is the mean transition frequency and $\Delta\bar{\varepsilon}_k = R_k\bar{\nu}_k/A\sqrt{2\pi}\,\sigma_k$ is the magnitude. A is a numerical constant (for more details see Section I of the Supplementary Information (SI)). For a given transition k, the rotational strength, R_k, depends on the relative orientation of the monomer electric dipole

Downloaded from https://academic.oup.com/nar/article/49/4/1872/6121468 by guest on 01 July 2021

transition moments, and is calculated from the diagonalization of the Hamiltonian that models the delocalized electronic states of the dApdA dinucleotide as a function of base stacking conformation. The Hamiltonian was formalized using the extended-dipole model (EDM). The parameters adopted in the EDM model are discussed in Section II of the SI. Specifically, Supplementary Figure S1 reports the angle that defines the direction of the electric dipole transition moment (EDTM) used in the CD calculations, while Supplementary Table S1 reports the experimental values for the magnitudes and the molecular frame orientations of the EDTMs used in our calculations. Supplementary Figure S2A shows the favorable comparison between the CD spectra calculated using the point dipole approximation (PDA) versus the extended dipole model (EDM) for the Watson–Crick B-form conformation. In Supplementary Figure S2B, we show a comparison between the CD spectra calculated using spectroscopic parameters for the Adenine monomer obtained from different experimental studies. For all of our CD calculations, we made use of the experimental parameters obtained by Williams *et al*. (7), which we list in Supplementary Table S2.

Markov state model analysis

The MD trajectories were analyzed using the Markov state model (MSM) PyEMMA software program (47). Briefly, we used the *k*-means++ algorithm (48,49) to construct a kinetically-relevant, balanced clustering of the trajectories (using the Euclidean criterion) by partitioning the 10^7 conformations into 100 initial microstates. A transition rate matrix was constructed for these microstates and then diagonalized into eigenvalues and eigenvectors. From the eigen-spectra of the transition probability matrix, we constructed five macrostates by implementing a minimum error propagation version of the Perron-cluster cluster analysis (PCCA+). We justified our choice for these five macrostates by considering the related conformational landscape and the implied interconversion time scales (see Section VI of the SI for further details). Rapidly interconverting molecular conformations were assigned to the same macrostate, while slowly interconverting conformations, which are separated by large barriers, occur between conformations that lie within different macrostates. By identifying and separating slowly interconverting conformations from rapidly interconverting ones, the MSM ensures that the slow processes obey Markovian statistics. To sample slow transitions, we adopted a lag-time of 500 ps, and confirmed that under these conditions Markovian behavior was satisfied by checking that the Chapman–Kolmogorov condition applies (50–52) (see Supplementary Figure S5 of the SI). We then calculated the CD spectrum for the configurations of dApdA that are contained in each macrostate.

RESULTS

Structural parameters of dinucleotides

As pointed out above, it has long been known that under physiological salt conditions the adjacent bases of each strand of duplex DNA in aqueous solution adopt helical conformations close to the Watson–Crick B-form, with an average inter-base separation $R \sim 3.5$ Å and a relative twist angle $\phi \sim 36°$ (see Figure 2 for parameter definitions). Spectroscopic studies of small oligonucleotides in solution have examined the various contributions to base stacking stability in duplex and ssDNA—i.e. the effects of hydrophobic bonding, backbone interactions, inter-base hydrogen bonding and cooperativity (8,18).

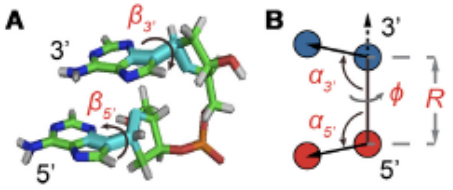

Figure 2. Structural coordinates for the dApdA dinucleotide monophosphate used in this work. (**A**) An atomic scale structure is shown with inter-base roll angles $\beta_{3'}$ and $\beta_{5'}$. (**B**) Virtual atoms are shown with blue and red spheres positioned within the planes of the 5′ and 3′ bases, respectively, with inter-base separation R, tilt angles $\alpha_{3'}$ and $\alpha_{5'}$, and dihedral twist ϕ (see SI for further details).

Free energy landscapes as a function of structural parameters and varying salt concentration

Prior studies of the dApdA dinucleotide monophosphate used CD spectroscopy to investigate changes in base conformation as a function of salt concentration, in order to elucidate the roles of the solvent ions in controlling dinucleotide conformation (8,18). These studies concluded that the predominant conformation for these truncated ssDNA molecules at physiological salt conditions is a stacked form close to the right-handed Watson–Crick B-form conformation, and that increasing the salt concentration appeared to destabilize this B-form conformation. As we discuss further below, the results of our analyses suggest that the dApdA system is, in fact, more accurately described as an equilibrium distribution of primarily three distinct stacked conformations.

We performed MD simulations of dApdA in aqueous solution at increasing salt concentrations with [NaCl] = 0.1, 0.5, 1.0, 1.05, 1.2 and 1.5 M, as described in the Materials and Methods (MM) section (see Figure 1 for a snapshot of a sample configuration). For each 2 μs simulation run, we constructed a histogram representing the probability $P(R, \phi)$ of finding a configuration at a given value of the structural parameters R and ϕ and the related free energy values, $G(R, \phi) = -k_B T \ln P(R, \phi)$ (see also Supplementary Figure S4, which reports the parameters adopted to calculate the FES and perform the MSM analysis, and related discussion in the SI). An example of such a free energy contour diagram plot is shown in Figure 3A. To ensure that the FES represents the system at equilibrium, we show, in Supplementary Figure S8 of the SI, the time autocorrelation function of the fluctuations in inter-base separation. Because the function is found to decay in ∼10 ns, which compares well with the 2 μs of simulation time, the system can be considered to be in equilibrium.

Two additional sets of orientational coordinates per base—the base tilt angles, $\alpha_{3'}$ and $\alpha_{5'}$, and the roll angles, $\beta_{3'}$ and $\beta_{5'}$—are needed to fully specify the dinucleotide con-

Downloaded from https://academic.oup.com/nar/article/49/4/1872/6121468 by guest on 01 July 2021

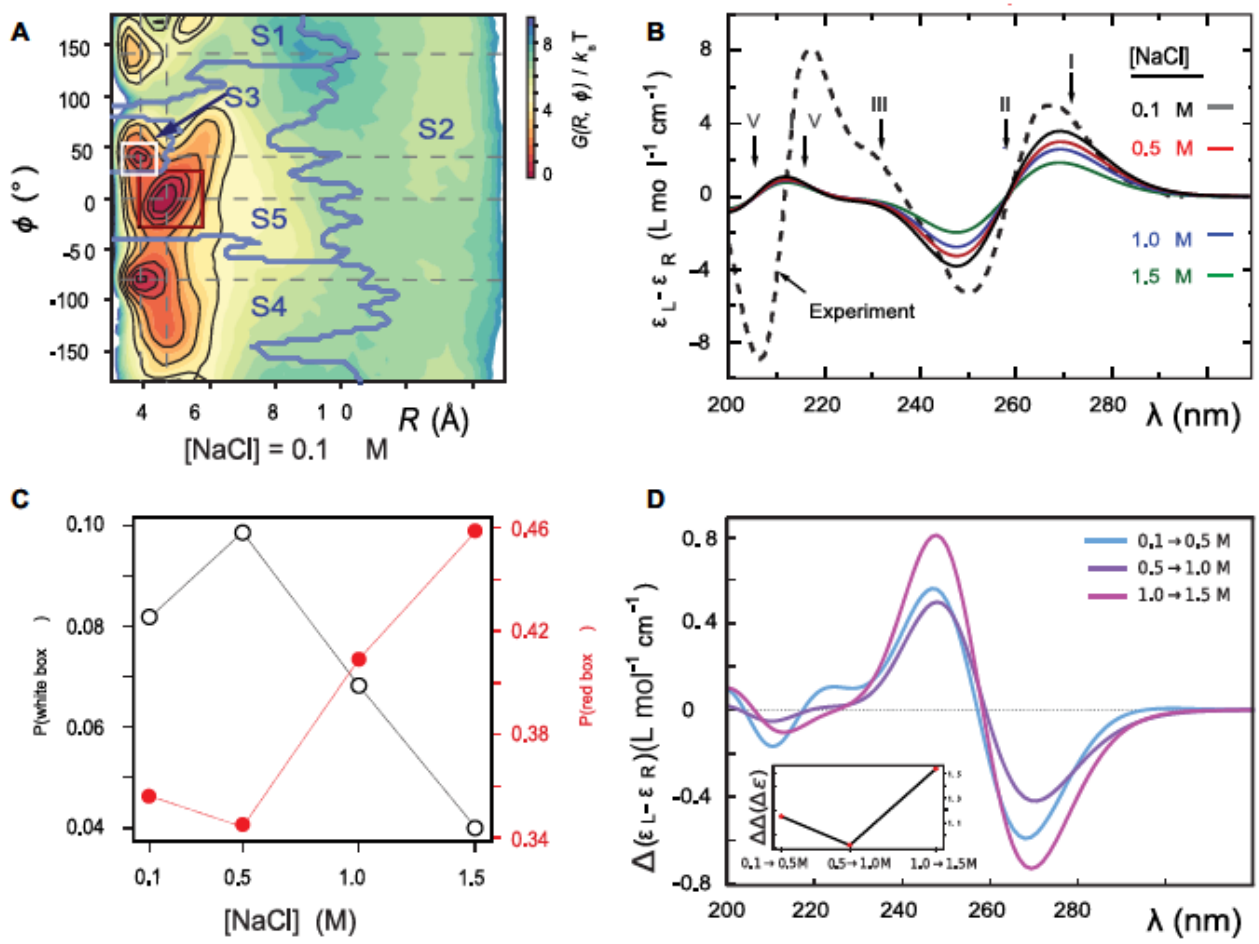

Figure 3. (**A**) Free energy landscape $G(R, \phi)$ as a function of the inter-base separation R and the dihedral twist angle ϕ, as obtained from 2 μs MD simulations of the dApdA dinucleotide at [NaCl] = 0.1 M. The coordinates corresponding to the canonical (average) B-form conformation ($R = 3.6$ Å and $\phi = 36°$) are included in the white (outlined) square, while the unstacked 'achiral' conformations are included in the red (outlined) square. The five macrostate regions, labeled S1–S5, were identified through the Markov State Modeling procedure (see related section). (**B**) CD spectra of dApdA were determined from 2 μs MD simulations at salt concentrations [NaCl] = 0.1 (black), 0.5 (red), 1.0 (blue) and 1.5 M (green). Differences between the calculated spectra are greater than the error bars (shown as the width of the colored lines), which were determined from the standard error of the mean from five block averages. The experimental CD spectrum (dashed black curve) was taken from (15). Roman numerals indicate the wavelengths of the electronic transitions of the uncoupled adenine monomers (see Supplementary Table S1 in the SI), which are used as input for our calculations. (**C**) The local probabilities of the B-like stacked conformation and the unstacked 'achiral' conformation was calculated as the sum of states contained within the boundaries defined by the white and red squares, respectively, shown in panel (A). These probabilities are shown as a function of salt concentration [NaCl] = 0.1, 0.5, 1.0 and 1.5 M. The relative population of stacked and unstacked conformations changes abruptly around 1 M concentration. (**D**) Differences between the CD spectra at increasing salt concentration. The difference between calculated CD spectra are shown for incremental increases of the salt concentration. The peak-to-peak amplitude of the difference CD spectra decreases dramatically when the concentration is raised above [NaCl] = 1 M. This is reflected by the abrupt change in the peak-to-peak amplitude of the difference CD spectra shown in the inset.

formation. However, our results indicate that the positions of the local minima in the FES depend largely on the inter-base separation R and the dihedral twist angle ϕ and are less sensitive to changes in the tilt and roll angles. While all of the above structural parameters are specified in our calculations of the CD spectra and of structural and dynamical distribution functions, the visual representation of the Free Energy Surface (FES) is conveniently reported as a function of R and ϕ. The FES $G(R, \phi)$ of the dApdA dinucleotide shown in Figure 3A applies to [NaCl] = 0.1 M, which is close to the monovalent salt concentration under physiological conditions. Using the same procedure, we also determined $G(R, \phi)$ at increasing salt concentrations (surfaces not shown). To test the validity of the FESs shown in Figure 3A, we used the molecular configurations obtained from our 2 μs MD trajectories to calculate the CD spectra for dApdA. The results, as a function of salt concentration and using the procedures described in the MM and the SI sections, are shown in Figure 3B.

For the lowest salt concentration, [NaCl] = 0.1 M, we compared our calculations to the experimental CD spectrum of the dApdA dinucleotide obtained under these same conditions (see Figure 3B) (15). We obtained good agreement between the experimental and calculated CD in the long wavelength region of the spectrum (240–300 nm). In principle one could achieve a quantitative agreement with

Downloaded from https://academic.oup.com/nar/article/49/4/1872/6121468 by guest on 01 July 2021

the experimental spectrum by optimizing some parameters in the calculation of the theoretical spectra, and thus obtain a better fit of the theoretical predictions to the experimental spectra. However, given the number of possible adjustable parameters, such a procedure would not provide any new information. We prefer, instead, to independently set the parameters in our calculations and then to discuss their predictions. We note that the agreement is less favorable in the short wavelength region (200–240 nm) of the spectrum, where the peak features are slightly blue-shifted and exhibit smaller amplitudes than the experiment. This latter disagreement is not surprising, given that the CD spectra at shorter wavelengths are strongly perturbed by the high density of nearly degenerate electronic states, which makes the theoretical methods we employ in our calculations less accurate in this wavelength range.

In general, we find that the positions of the local minima within the free energy surfaces do not change with salt concentration, while their relative stabilities and equilibrium distributions do depend on this variable. The FES in Figure 3A shows that the dApdA dinucleotide exists primarily as a mixture of the two chiral conformations with opposite handedness ($\phi = 40°$ and $-80°$) and nearly stacked inter-base separation $R = 3.8$ Å, together with an achiral conformation that shows no stacking of the bases ($\phi = 0°$) and a significantly larger inter-base separation $R = 4.7$ Å. Henceforth, we will designate as 'chiral' a conformation that exhibits chiral stacking of the bases, and as 'achiral' conformations with no stacking of the bases, even though some components of the molecule, like the sugar, do of course retain their 'chemical chirality.'

To study the effects of increasing salt concentration on the population of the chiral and achiral conformational states, we report in Figure 3C the local probabilities calculated as the sum of the states contained within the areas of the FES defined by the red and white squares (panel *A*), respectively, for the chiral state with coordinates (3.8 Å, 40°) and for the achiral state with coordinates (4.7 Å, 0°) as a function of the salt concentration. We note that as the salt concentration is increased to [NaCl] = 0.5 M, the local probability of the chiral state with coordinates (3.8 Å, 40°) increases, while the weight of the achiral state slightly decreases. A further increase of the salt concentration to [NaCl] ∼ 1 M begins to destabilize both of the stacked conformations in favor of the unstacked one, with the weight of the achiral state with coordinates (4.7 Å, 0°) increasing strongly. We observe a similar dependence on the salt concentration for the CD spectrum, which depends on the distribution of stacked bases. In Figure 3D, we plot the difference CD spectrum for incremental changes of the salt concentration. For incremental increases of the salt concentration below [NaCl] = 1 M (0.1 → 0.5 M, 0.5 → 1.0 M), the difference CD spectrum shows little variation. However, for the incremental increase of 1.0 → 1.5 M, the difference CD spectrum undergoes a pronounced change. This change is also reflected by the value of the peak-to-peak amplitude of the difference CD spectrum (i.e. the difference between the positive peak value at 245 nm and the negative peak value at 270 nm), which is shown in the inset of Figure 3D. Note that these findings are in agreement with the salt-dependent changes in the CD amplitude of this dinucleotide reported in the work of Johnson and Schleich (18).

The above findings are in qualitative agreement with experiments involving the thermal melting of duplex DNA structures in NaCl, where increases in the concentration of monovalent ions tend to first stabilize the stacked conformation, resulting in an increase in the melting temperature. Then, at higher salt concentration (around [NaCl] = 1 M) this trend reverses, and the further addition of counterions slightly decreases the stability of the dsDNA conformation (39–41). For duplex DNA, this behavior is generally explained by assuming that an increase in salt concentration facilitates the screening of the negative charges situated on the phosphates in the DNA backbone, rendering the backbone more stable. However, at monovalent salt concentrations around 1 M, the concentration of ions in solutions becomes equivalent to the concentration of counterions closely bound to the phosphate backbone under ion condensation conditions. As a consequence, additional increases in salt concentration cannot further stabilize the double helix and other mechanisms (presumably 'Hofmeister effects' (19,53–56) come into play. Mechanisms involving the stabilization of long duplex DNA molecules by screening the repulsion between backbone phosphates cannot apply to dApdA, since only one phosphate is present. However, the counterions can alter the relative stabilities of the various conformations available to the dApdA dinucleotide by effectively neutralizing the negative charge of the single phosphate.

Distributions of ions and water molecules around dinucleotides

To examine the roles of salt concentration on the observed structural transition we used the results of our MD simulations to calculate the distributions of the ions and water molecules of the solvent environment in the immediate vicinity of dApdA. This study provides physical insights into the origins of the changes in equilibrium base stacking conformations of this dinucleotide with increasing salt concentration (1–4,15).

The radial distribution function (RDF) of species j around species i is defined:

$$g_{i-j}(r) = \frac{\langle \rho_j(r) \rangle}{\langle \rho_j \rangle} \quad (1)$$

Figure 4A and B shows the RDFs of the dApdA system at the lowest and highest salt concentrations we examined; i.e. [NaCl] = 0.1 and 1.5 M, respectively. The position-dependent oscillations of the RDFs reflect the local solvation shells of the water hydrogen atoms and of the ionic species relative to the central phosphate. At salt concentrations close to physiological conditions, ([NaCl] = 0.1 M, Figure 4A), the phosphate is coordinated with concentric ion shells, with the water hydrogen atoms forming interstitial layers between the shells. The RDF for water hydrogen atoms appears to be independent of salt concentration, with its first peak centered at $r = 2.8$ Å and its second peak at $r = 4.2$ Å. The RDFs for sodium and chloride ions, on the other hand, oscillate at half the spatial frequency of that of

Downloaded from https://academic.oup.com/nar/article/49/4/1872/6121468 by guest on 01 July 2021

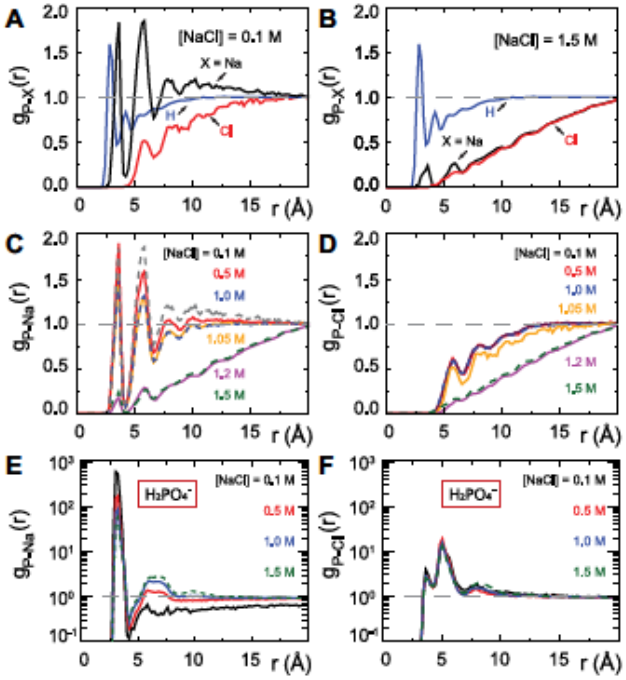

Figure 4. Radial distribution functions (RDFs) [Equation (1)] obtained from MD simulations of dApdA between Na^+, Cl^-, and the H atoms of water and the P atom of the anionic phosphate of the dApdA dinucleotide at salt concentrations ([NaCl]) of (**A**) 0.1 M and (**B**) 1.5 M. RDFs for (**C**) sodium ions and (**D**) chloride ions over the range of salt concentrations [NaCl] = 0.1, 0.5, 1.0, 1.05, 1.2 and 1.5 M. RDFs for sodium (**E**) and chloride (**F**) ions obtained from MD simulations of the phosphate anion $H_2PO_4^-$ at the salt concentrations ([NaCl = 0.1 (black), 0.5 (red), 1.0 (blue), 1.5 M (light blue). Unlike the RDF plots for dApdA, the RDFs of $H_2PO_4^-$ in aqueous solutions do not show the sharp change in the ion shell structure at [NaCl] ≅ 1.0 M (see also text).

the water hydrogen atoms. The RDF for sodium ions has its first peak at $r = 3.6$ Å, which coincides with a trough for the water hydrogen atoms at this distance. Similarly, a trough for sodium ions occurs at $r = 4.2$ Å, which coincides with the second hydration shell for the water hydrogen atoms. The first ion shell for chloride ions occurs at $r = 5.8$ Å, which is the same position as the second ion shell for sodium ions. In general, the nth chloride ion shell occurs at approximately the same position as the $(n + 1)$th sodium ion shell, indicating that these ion shells have mixed compositions. As shown in Figure 4B, the relatively well-defined boundaries between successive ion shells seen at the lowest salt concentrations become diffuse at the highest salt concentration tested ([NaCl] = 1.5 M).

Our observation of a well-ordered structure of successive ion shells at low salt concentration is largely consistent with simple models of counterion condensation, which is an important contributing factor to the stability of larger nucleic acid molecules (57,58). Figure 4C and D shows, respectively, the RDFs of sodium and chloride ions, each as a function of salt concentration. For both ions, the RDFs appear to change little over salt concentrations between [NaCl] = 0.1–1.0 M, yet exhibit an abrupt loss of ion shell structure at salt concentrations slightly greater than 1.0 M.

To illuminate the role(s) of the adenine bases in this situation, we performed a set of 400 ns simulations of $H_2PO_4^-$ at increasing monovalent salt concentration ([NaCl] = 0.1, 0.5, 1.0 and 1.5 M), and studied the ion distributions around a singly-charged phosphate ion, $H_2PO_4^-$, in aqueous solution (see Figure 4E and F). In $H_2PO_4^-$ we observed an alternating structure of positive and negative ion shells consistent with simple models of counterion condensation. However, we found no signature of the abrupt disruption of the ion shell structure at salt concentrations greater than 1.0 M that was observed with the dApdA dinucleotide.

We next turned our attention to a closer examination of the solvent orientation around dApdA. As mentioned previously, the structure of the water, which is reported as the position-dependent RDF of the water hydrogen atoms relative to P, $g_{P-H}(r)$, does not change significantly with salt concentration (see Figure 4A and B). More detailed behavior is observed in the position-dependent orientational distribution function (ODF) of the water dipole moment as a function of its separation from the central P atom. The ODF is defined as the average cosine, $\langle \cos\theta(r) \rangle$, of the angle θ that subtends the permanent dipole moment of the water molecule, $\vec{\mu}_{H_2O}$, and the vector connecting the P atom to the water O atom, $\overrightarrow{PO}_{H_2O}$, as shown in Figure 5A.

Figure 5B shows the ODFs of water relative to the central phosphate of dApdA as a function of salt concentration. It also shows the RDF of the water hydrogen atoms. The position dependence of the ODFs shown in Figure 5B exhibits damped oscillations that vary across successive hydration layers for all salt concentrations, similar to the behavior observed for the ion shell structures shown in Figure 4. The ODFs show a sharply pronounced feature centered at $r = 2.8$ Å, which is coincident with the first peak of the RDF. The shapes of the underlying distributions of the angle θ within a narrow range of distances r ensures that $\langle \cos\theta \rangle \approx \cos\langle\theta\rangle$ (orange points in Figure 5B). The distributions of the angle θ for a given hydration shell, with each distribution corresponding to one orange point in Figure 5B, are reported in Supplementary Figure S7 of the SI. Thus, the narrow feature at $r = 2.8$ Å has an approximate peak value of $\cos\langle\theta\rangle = -0.8$, which indicates that the water H atoms within this first hydration shell are highly oriented with dipole moment $\vec{\mu}_{H_2O}$ directed toward the central P. Furthermore, the presence of the broadened shoulder centered near the second hydration layer (at $r = 4.2$ Å), with peak value approximately $\cos\langle\theta\rangle = -0.6$, indicates the preferential orientation of the O-H bond vectors of water molecules within the second hydration shell toward the oxygens of water molecules within the first hydration shell. We thus see that hydrogen bonding interactions between water molecules of the first and second hydration shells are stronger than the Coulomb interaction between the negatively charged phosphate and the water dipole moments of the second hydration shell. We further note that the distribution of angles θ over a given range of distances r broadens nonuniformly as the distance from the central P increases, indicating the presence of hydrogen bonding between successive hydration layers and the ensuing loss of orientational correlation between the water dipoles and the central P. At the separation $r = 5$ Å, the values of the ODFs

Downloaded from https://academic.oup.com/nar/article/49/4/1872/6121468 by guest on 01 July 2021

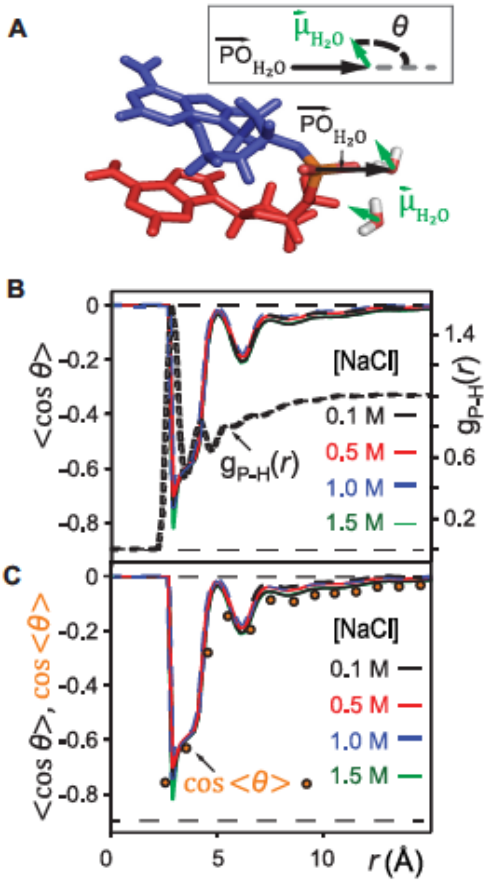

Figure 5. (**A**) Definition of the angle θ, which subtends the permanent dipole moment of the water molecule $\vec{\mu}_{H_2O}$ and the vector $\overrightarrow{PO}_{H_2O}$ connecting the phosphorous atom to the oxygen atom of the water molecule. (**B**) Orientation distribution functions (ODFs) for the dipole of the water molecule relative to the phosphate–oxygen (water) bond and RDF of the hydrogen of water, $g(r)_{P-H}$, of Figure 4A. (**C**) Superimposed on the ODFs defined as $< \cos\theta >$, are the orange points indicating the cosine of the average angle, $cos\langle\theta\rangle$.

are approximately zero, indicating the absence of orientational alignment. A recurrence of partial orientational order occurs at separation $r = 6$ Å, which appears to coincide approximately with the position of the first ion shell of the Cl^- ions.

We note that the ODF exhibits a weak, but clear, dependence on the salt concentration. For the case of [NaCl] = 0.1 M, the sharp feature at $r = 2.8$ Å indicates a pronounced orientation, which becomes slightly less ordered for [NaCl] = 0.5 M. At the higher salt concentrations of [NaCl] = 1.0 and 1.5 M, the orientation of the water dipole moments become slightly more ordered. The changes in the ODF of the water molecules as a function of increasing ion concentration are small. Rather, the leading factor in determining the stabilities of the conformations of the dinucleotide structure in solution appears to involve the distribution of monovalent ions, and the modification of this distribution with increasing ion concentrations (see Figure 4A–D). Water, however, does appear to play a role through its orientation, which is both distance and weakly salt-concentration-dependent. Interestingly, this study also shows that the stabilization of dApdA stacking by increasing counterion concentrations, and the observed sharp transition of the ion structure around 1 M are dependent on the presence of the bases of the dApdA dinucleotide, and do not occur when the ionized phosphate molecule is present alone (Figure 4E – F). The consistency of the observed trend with the effects of increasing salt concentration on the experimental melting curves of DNA suggests that this ion-related base stacking mechanism of DNA stabilization is already present and operational, even at the level of the isolated dinucleotide.

Markov state model analysis of the free energy landscapes of the dApdA dinucleotide and comparison with CD spectral analysis

The theoretical representation of the CD spectrum for a flexible molecule in solution is the summation of contributions from the myriad microscopic conformational states (i.e. microstates) that exist at equilibrium. Intuitively, we expect the dApdA dinucleotide to fluctuate between various 'open' and 'closed' base conformations, which in turn are stabilized (or destabilized) by the surrounding hydration and ion shells. We first determined the CD spectrum by summing over equally weighted contributions from the 10 million microstates that are sampled from each of our MD simulations (see Figure 3B). Although the above 'brute-force' approach is straightforward, it suffers from two significant limitations: (i) it provides little insight into the interpretation of CD in terms of specific molecular conformations and (ii) it becomes computationally inefficient if one adopts more sophisticated quantum chemical models to calculate the CD spectrum beyond the extended dipole model used here, because one would need to perform advanced calculations for each of the 10 million microstates. In reality, only a relatively small subset of the total number of possible conformational states is expected to contribute significantly to the measured CD spectrum. The specific states that dominate the CD are the stacked and chiral conformations of the dinucleotide, for which both the electronic coupling between monomer electric dipole transition moments (or EDTMs) and the rotational strengths resulting from these couplings are significant (see SI). Conformational states that are unstacked, in addition to those that are stacked and essentially achiral, contribute much less to the CD spectrum.

To determine the dApdA configurations that are most relevant for the interpretation of the CD spectra, we used a Markov state model (MSM) analysis (20–22) to subdivide the 10 million microstates obtained from our MD simulations into a relatively small number (five) of 'macrostates', each of which is associated with a distinctive region of the free energy landscape (see Figure 6A). Each macrostate represents a collection of conformationally related states, or 'microstates', which rapidly interconvert during the simulation, while slow transitions between macrostates require crossing large energy barriers. Starting from the MSM analysis, we calculated the CD spectrum as the sum of those, unequally weighted, macrostates.

Thus, the kinetic processes that occur in the simulations are partitioned between those that occur faster than the 'lag

Downloaded from https://academic.oup.com/nar/article/49/4/1872/6121468 by guest on 01 July 2021

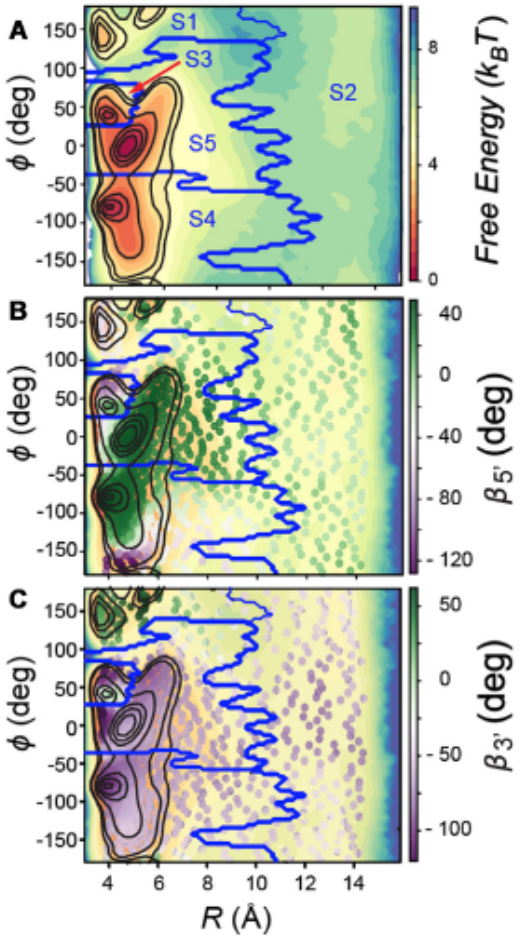

Figure 6. (**A**) The free energy landscape $G(R, \phi)$ of the dApdA dinucleotide (shown in Figure 2A) is sub-divided by dark blue boundaries into five regions (labeled S1 – S5), which are called 'macrostates.' The macrostate assignments were derived by performing a Markov state model (MSM) analysis of MD simulation data for [NaCl] = 0.1 M. The *anti* (Watson–Crick) form conformation is contained within the boundaries of the S3 macrostate, while the *syn* (Hoogsteen) containing form is included within the boundaries of the S4 macrostate. Superimposed on the free energy landscape $G(R, \phi)$ of the dApdA dinucleotide we show the orientation of the 5′ base (**B**) and of the 3′ base (**C**), respectively. The macrostate S3, which contains the *anti* form conformation, correctly displays both bases with positive orientation (green free energy minimum of S3 in both panels *B* and *C*), while macrostates S4 and S5, which contain a *syn* base presents the 3′ base flipped with respect to the 5′ base (green in panel *B* and purple in panel *C*).

time' (in this study, $\tau = 500$ ps) and those that occur more slowly than this time scale. Transitions between conformations within a given macrostate occur frequently and are non-Markovian, while transitions between conformations belonging to different macrostates occur less frequently and are Markovian. The 'lag time' is defined as the time that fulfills the above-stated condition of Markovian transitions between conformations belonging to different macrostates (for details see SI and, in particular, Supplementary Figure S5, which tests the Chapman–Kolmogorov condition, thus ensuring the Markovian nature of our partitioning of the FES into five states. Supplementary Table S3 shows that the MSM analysis is insensitive to the choice of the number of microstates).

To further reduce the computational requirements for the calculation of the CD spectrum, we identified one averaged structure, together with its relative weight, for each of the five key macrostates that are relevant to the CD observable. We found that the total CD spectrum can be accurately represented by the weighted sum of the contributions from these five averaged structures (see Supplementary Figure S6A–F in the SI), which could be used for modeling of the CD spectrum using more advanced quantum chemical models.

In Figure 6A, we show the free energy landscape for the dApdA dinucleotide monophosphate in 0.1 M salt (NaCl), and its subdivision into five macrostates (indicated by dark blue contour lines and labeled S1–S5), which we established using our MSM analysis approach. Each of the five macrostates exhibits qualitatively different behavior in terms of the relative stabilities of the dinucleotide conformation. The S1 macrostate includes 264,608 microstate configurations (2.6% of the total 10 million) and is dominated by a relatively shallow free energy basin with a narrow range of values for the inter-base separation $R < 6$ Å and relative twist angle: $100° < \phi < 180°$. The S2 macrostate, on the other hand, describes a relatively broad and featureless region of the free energy landscape, which encompasses a wide range of 'open' and 'unstacked' values for the inter-base separation ($R > 10$ Å) and unrestricted twist angle: $-180° < \phi < 180°$. Like the S1 macrostate, the S2 macrostate represents a minority of the total population, with just 276 786 microstates (2.8% of the total 10 million).

The majority of the total conformation population is contained in the combined S3, S4 and S5 macrostates, with the number of microstates in S3: 895 636 (9.0%); in S4: 3 729 206 (37.3%); and in S5: 4 833 758 (48.3%). Moreover, the S3 macrostate contains the free energy minimum with $R = 3.8$ Å and $\phi = 40°$, the S4 macrostate contains a minimum with $R = 3.8$ Å and $\phi = -80°$, and the S5 macrostate contains a minimum with $R = 4.7$ Å and $\phi = 0°$. We thus identify the S3 macrostate with an ensemble of stacked right-handed chiral conformations that include the *anti* form; the S4 macrostate with an ensemble of stacked left-handed chiral conformations, including the precursor of the Hoogsteen structure; and the S5 macrostate with one slightly less stacked and more achiral conformation, which also includes a *syn* structure (30–37). The borders between macrostates show a 'fine structure' that represents the maximum of the energy at the top of the free energy barriers, where the states are less frequently sampled by the simulation. Thus, the high energy regions in the free energy map may display roughness, which can be smoothed to avoid overfitting (59,60). However, this step is not needed in our study because the results of our analysis depend largely on the minima of the free energy maps, which are statistically well sampled.

To confirm the presence of Hoogsteen-like structures in the S4 and S5 regions, we present—in Figure 6B and C—a study of the roll angles $\beta_{5'}$ and $\beta_{3'}$, for the 5′ and 3′ base, respectively. It is known, for structures containing a Hoogsteen conformation, that one of the two bases in the dApdA dinucleotide is 'flipped' relative to the 'standard' conformation characteristic of the Watson–Crick geometry. Fig-

Downloaded from https://academic.oup.com/nar/article/49/4/1872/6121468 by guest on 01 July 2021

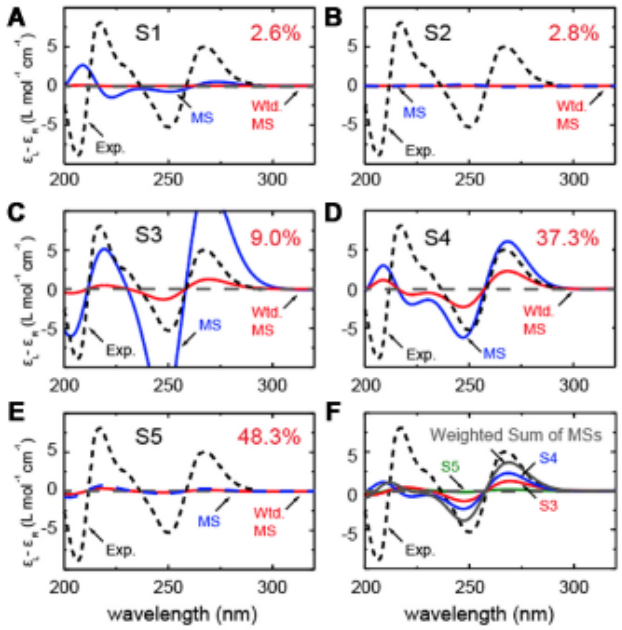

Figure 7. Macrostate decomposition of the CD spectrum of the dApdA dinucleotide by Markov state model (MSM) analysis of MD simulation data for [NaCl] = 0.1 M. The total CD spectrum is calculated from 10 million MD frames (or microstates), and the component spectra for macrostates (**A**) S1, (**B**) S2, (**C**) S3, (**D**) S4 and (**E**) S5 constitute 2.6, 2.8, 9.0, 37.3 and 48.3% of the total CD spectrum, respectively. The component CD spectra for each macrostate are shown in blue, and the number-fraction weighted contributions are shown in red. (**F**) The sum of number-fraction weighted macrostate contributions to the total CD is shown in gray. Also shown separately are the number-weighted contributions of macrostates S3, S4 and S5 (green, blue and red, respectively). In all panels, the experimental CD spectrum (from (12)) is shown as dashed black curves.

ure 6B shows that in the S3 macrostate the most stable structures have a positive roll angle $\beta_{5'}$ for the 5′ base (green). Figure 6C shows, instead, that while the roll angle for the 3′ base is still positive in the microstate S3, the same 3′ base is flipped in microstates S4 and S5 (purple), confirming the presence of a *syn* Hoogsteen-like conformation in macrostates S4 and S5, and the *anti* Watson–Crick-like form in macrostate S3.

In Figure 7A–E, we compare the experimental CD spectrum (dashed black curve) to our CD calculations corresponding to each of the five macrostates (blue curves), which are based on summing over the microstate configurations that lie within the partitioned boundaries of the free energy landscape shown in Figure 6A. We also show the proportionately weighted contribution of each macrostate to the CD spectrum (red). From Figure 7A and B, we see that the S1 and S2 macrostates, which represent minority fractions of the total population (2.6% and 2.8%, respectively), do not contribute significantly to the CD spectrum.

Similarly, the S3 macrostate (Figure 7C), which contains the coordinates of the B-form conformation, also represents a comparably small fraction of the total population (9.0%). On the other hand, the S4 macrostate (Figure 7D) contains a significant fraction of the total population (37.3%) and is largely composed of left-handed base-stacked conformations, which gives rise to a strong CD signal. We note that the calculated CD spectrum of the S4 macrostate has a similar 'right-handed' shape (in the long wavelength regime) to that of the S3 macrostate, in spite of their apparent opposite chiral symmetries. This is consistent with the *syn* structure (i.e. with roll angles $\beta_{3'} \approx 180°$ and $\beta_{5'} \approx 0°$). The detailed calculation of the spectra for all five macrostates is reported in Supplementary Figure S3 of the SI, which shows the spectral decomposition of the degenerate CD spectrum for the average structure of each macrostate.

From the spectral decomposition it is straightforward to see that the flipping of one base is responsible for a CD spectrum that is consistent in the Watson–Crick structure and in the *syn* conformation of dApdA. We note that this behavior may not be observed in dinucleotides with different base compositions, because the transition dipoles are different. Although the S5 macrostate (Figure 7E) represents the highest fraction of the total population (48.3%), it is dominated by an achiral and slightly unstacked *syn* conformation which, because of its symmetry, results in a negligible CD contribution to the total spectrum. In Figure 7F, we show the individual weighted contributions for the S3 (9.0%), S4 (37.3%) and S5 (48.3%) macrostates, in addition to the weighted sum of all the macrostates (gray curve). We thus see that the favorable agreement we observe between experiment and theory in the long wavelength regime is essentially the result of two significant contributions, a minor contribution from macrostate S3 and a larger contribution from macrostate S4.

Having identified the key macrostates relevant to the CD observable, we used this information to determine the smallest number of structural parameters necessary to characterize these macrostates. We thus identified five averaged conformations, one for each macrostate, which, properly weighted, were used to calculate the CD spectrum. The comparison between the contribution to the CD spectrum from all the conformational states in a macrostate and the contribution from the averaged macrostate structure are shown in Supplementary Figure S6 of the SI, with structural parameters listed in Supplementary Table S6. The calculation of the CD spectrum with only five conformations is in good agreement with the complete calculation, while it greatly speeds up the computation time needed to calculate the CD spectrum. In principle, such structural models can be used for the general interpretation of any spectroscopic measurement performed on the dApdA system.

In reconsidering the previous interpretations of the CD spectrum by Lowe and Schellman, and given that the signal from the un-stacked mono-nucleotide is comparatively negligible, our study suggests that the stacked 'native' form of the dinucleotide is primarily given by the sum of the S3 and S4 states, because the S1 state is less densely populated. Analogously, the unstacked 'denaturate' state corresponds in this study to the S5 state, which is more populated than the unstacked S2 state.

The large degree of conformational disorder that characterizes macrostate S2 contrasts with the highly ordered macrostates S3 and S4. The stabilities of macrostates S3 and S4, relative to macrostate S5, are reminiscent of the 'solvophobic' models for nucleic acid base stacking (2,11,12), in which the 'stacked' macrostates S3 and S4 are favored due to enthalpic base stacking interactions (61), which

Downloaded from https://academic.oup.com/nar/article/49/4/1872/6121468 by guest on 01 July 2021

offsets the configurational entropy of the disordered S2 macrostate. Solvophobic base stacking is known to be favored by a decrease in the enthalpy ΔH and opposed by a decrease in the entropy ΔS. Solvophobic bonding, as defined here, is 'enthalpically driven' and differs significantly from hydrophobic bonding, which is generally thought to drive protein folding (62) by a positive change in the entropy of the system. Such physical models are supported by studies that examine the stabilizing and destabilizing effects on base stacking by various salts and other solvent additives (10,18).

Mean first passage times (MFPTs) for dApdA macrostates and pathways of macrostate interconversion

While CD spectra provide a useful measure of the stationary (equilibrium) properties of the dApdA system, they do not provide information about the dynamic processes involved in state-to-state interconversion. In this section we apply the results of our MSM analysis of MD trajectories to the investigation of the kinetic pathways associated with the free energy landscape, and to identify pathways of interconversion between the various stacked and unstacked macrostates.

To characterize the kinetic properties of the dApdA system, we examined the mean first passage times (MFPTs) of the five macrostates, which are assigned to the regions of the free energy landscape shown in Figure 6A. The MFPT $\tau_{i\to f}$ is the average time for the system to undergo a transition to state f, provided that it was initially in state i (21,63). We determined the MFPTs for the free energy landscape of dApdA at salt concentration [NaCl] = 0.1 M (see full data set in Supplementary Table S4 of the SI. Also, Supplementary Table S5 shows that the MFPTs are insensitive to the number of microstates selected in the MSM analysis). Macrostate S2 represents the region of the free energy landscape with the greatest degree of conformational disorder; thus, it can be considered to serve as an end-state for base-unstacking.

Moreover, while macrostate S3 is approximately B-form in character, the relative roll angles of macrostates S4 and S5 are >90°, which in each case corresponds to a base configuration that has been flipped into the Hoogsteen-like conformation. Thus, the process of 'base-flipping' may play an important role in the dynamics of the dApdA system, although in longer strands of (especially) duplex DNA, such flipping may be suppressed by the overall cooperativity that controls the order-disorder transitions for these larger macromolecular species. Nevertheless, these studies of the less cooperatively stabilized dinucleotide may provide insight into structural rearrangements that in principle could, and likely—with some frequency—do, occur in larger biologically relevant DNA macromolecules.

We used the transition path theory (TPT) method (25–29) to determine the frequency of events in which an initially base-stacked macrostate (e.g. S3 or S4) undergoes successive conformational changes that permit entry into the region of the free energy landscape characterized by the 'final' unstacked macrostate S2. When the system initially occupies macrostate S3, which corresponds to the average base stacking of the Watson–Crick B-form, we found that the dominant pathway leading to macrostate S2 (with 46% probability) was S3 → S5 → S2. Thus, base-unstacking from the right-handed B-form conformation occurs predominantly by a two-step process through the achiral S5 intermediate, in which one of the adenine bases has been flipped. The remaining, less prevalent base-unstacking pathways were S3 → S4 → S2 (with 26% probability); S3 → S5 → S4 → S2 (with 15% probability); and S3 → S2 (with 10% probability). When the system occupied initially the left-handed and base-flipped macrostate S4, which corresponds to a Hoogsteen base-stacking configuration, the two most prevalent unstacking pathways were the one-step S4 → S2 pathway (with 47% probability) and the two-step S4 → S5 → S2 pathway (with 40% probability).

We see that, in general, transitions to the most sparsely populated macrostates S1 and S2 occur relatively slowly (in ∼35–60 ns), while transitions to the most highly populated macrostate S5 occur relatively quickly (in ∼2–5 ns), suggesting that the macrostate S5 acts as a common intermediate for the pathways between the other macrostates for the stacking-unstacking transition.

Because the energy barriers in dApdA are small, the height(s) of the barrier(s) that the system has to overcome to transition between any two macrostates is close to the difference in free energy between the two states. Thus, the kinetics of the interconversion between macrostates are driven primarily by their relative stabilities. It is reasonable to expect that cooperativity in base stacking will increase the heights of the energy barriers between conformational states in both the ssDNA and the dsDNA. Such information can provide new insights into the mechanisms of base stacking-unstacking transitions in nucleic acids and the possible role of these processes in biologically important protein-nucleic acid interactions.

DISCUSSION

Structural and dynamic characterization of 'breathing' fluctuations at the dinucleotide level

Thermally activated breathing fluctuations, in which flanking nucleic acid bases spontaneously move away from their stacked and hydrogen-bonded conformations, are thought to be important initial steps in the pathways that lead to DNA denaturation and the specific binding of proteins to DNA (1–6). Despite their relevance, the details of the interactions and kinetics that control breathing fluctuations are still largely not understood. It is known, however, that the stacking interactions of the bases within nucleic acids are the dominant stabilizing forces of the native conformations that oppose the melting of DNA, while inter-strand base-base hydrogen bonding and cooperativity play less important stabilizing roles (10,17,61). Traditionally the equilibrium between stacked and unstacked base conformations has been studied by circular dichroism (CD) experiments, which are sensitive to the conformational chirality of the base stacking (15). Such measurements, however, are limited in the amount of information that they can provide because CD spectra cannot be directly inverted to determine the conformations that contribute to these spectroscopic signals.

CD spectroscopy is an important biophysical tool for the analysis of nucleic acid structure, in that the relationship between CD spectra and local nucleic acid base conformation

Downloaded from https://academic.oup.com/nar/article/49/4/1872/6121468 by guest on 01 July 2021

can be understood in terms of quantum chemical principles. Nevertheless, for many of these systems the free energy landscape can favor the simultaneous presence of multiple conformations at equilibrium, many of which may interconvert due to thermal fluctuations. Thus, the complexity of the free energy landscapes of nucleic acid systems is a significant obstacle for achieving a meaningful interpretation of CD spectra.

Solvophobic effects on the conformational stability of dinucleotides

Early studies by Lowe and Schellman of the base stacking-unstacking equilibrium focused on the CD spectrum of the dApdA dinucleotide monophosphate as a function of increasing monovalent salt concentration, because the stacking interactions of the elementary dinucleotide unit could be isolated and studied independently of other stabilizing factors (10). These studies concluded that the stacking-unstacking equilibrium of dinucleotides can be modeled as a two state transition, where the driving force for the stacking of the bases is 'solvophobic' in nature; i.e. driven by a decrease in the enthalpy of the process ($\Delta H \approx -6.6$ kcal mol^{-1} at $T = 293$K), and opposed by a decrease in the entropy of the system ($\Delta S = -23$ e.u. such that $-T\Delta S \approx 6.7$ kcal mol^{-1} at $T = 293$ K) (10,14,15,61). Thus, these workers concluded that the transition as a whole was likely driven – to a major extent—by rearrangements of the molecules of the solvent environment present (here water molecules and ionic species). However, these studies could not exclude the possibility that more than two states might contribute to the overall CD signal, and thus could not define the precise nature of the underlying conformations (10). They did determine, however, that each of the two states of the dApdA dinucleotide that contributed to the CD signal was most likely present as a number of similar configurations, and that the state with highest disorder and entropy was likely to be more stable at high temperatures and at higher monovalent salt concentrations.

CONCLUSIONS AND OVERVIEW

In the present study, we have established a methodology that can be used to relate the CD spectrum to the underlying relevant molecular conformations. We combined extensive MD simulations (μs in duration) with direct calculations of the CD spectrum. Our CD calculations were based on standard methods (24) and an extended-dipole model (EDM) (64) to estimate the exciton coupling between the electric dipole transition moments (EDTMs) of the adenine bases of dApdA. The EDM takes into account the finite length of the electronic transition charge distribution across the adenine chromophore, and it correctly describes the dependence of the electronic coupling on the inter-base twist angle ϕ and the relative tilt angle $\alpha_{5'} - \alpha_{3'}$ (coordinates defined in Figure 2). By calculating the CD spectrum for each of the 10 million conformations in the MD simulations, we obtained good agreement between our CD calculations and previously published experimental spectra of the dApdA system at approximately physiological salt concentration [NaCl] = 0.1 M (10,15). Nevertheless, the calculation of the CD spectrum by these procedures provided little insight into the important conformational states contributing to the CD spectrum, and can become computationally too expensive if sophisticated quantum chemical calculations are adopted to calculate the exciton couplings from the detailed electronic structure of the adenine bases.

To surmount this problem, we performed a Markov state modeling analysis of the free energy landscape of the dApdA dinucleotide and identified five kinetically separable macrostates, each containing conformational species that can rapidly interconvert. We then calculated a single averaged conformation to represent each of the five MSM macrostates, and we found that the total CD spectrum can be represented accurately by the weighted sum of the contributions from the averaged structures of these macrostates.

We found that only two states exhibit both stacked and chiral conformations, which are necessary to provide significant exciton coupling between monomer EDTMs and rotational strengths, thus contributing to the CD observable. The two states are conformational ensembles with opposite chirality, which contain the *anti* (Watson–Crick B) form (S3) and a *syn* (Hoogsteen) flipped-base conformation (S4), respectively. A third highly populated state is an achiral *syn* state, with a slightly unstacked conformation (S5) that does not contribute significantly to the CD signal. We observed that both the S3 and the S4 states provide right-handed CD features in the long wavelength region of the dApdA spectrum. These results are qualitatively consistent with the early hypothesis that two leading states dominate the CD spectrum, but now provide more detailed information about the nature of those states (10). We conclude that both the S3 and the S4 states contribute to the stacked conformation detected by Lowe and Schellman, while their unstacked conformation likely comprises the S5 state, which is the most populated, and to a lesser extent the fully unstacked state S2. Furthermore, our study shows that the Hoogsteen structure plays a key role in the mechanism of the stacking and unstacking pathways of the bases in dApdA, and possibly in DNA as well, as it is present in the highly populated stacked S4 and unstacked S5 conformations at all salt concentrations.

By connecting the CD spectrum of the dApdA dinucleotide to five leading conformational states as a function of salt concentration, we were able to obtain information about how the distribution of ion shell structures affects local base-stacking interactions. In agreement with early experiments by Johnson and Schleich (18), we observed that the effect of increasing salt was to decrease the magnitude of the CD signal over the 240–300 nm regime. In the CD experiments by Lowe and Schellman (10), the decrease in CD signal was accompanied by a shift in the equilibrium population of open (unstacked) achiral conformations relative to closed chiral conformations. Our findings show an initial increase of the base stacking stability with increasing monovalent salt concentration (NaCl or KCl), followed by a decrease of stability at salt concentrations higher than 1.0 M. By analyzing radial distribution functions and orientation distribution functions of the ions and water solvent, respectively, we observed that the changes in local base stacking conformation at high salt concentrations are correlated strongly with the disruption of the ion shell boundaries, and

Downloaded from https://academic.oup.com/nar/article/49/4/1872/6121468 by guest on 01 July 2021

weakly with a change in the orientations of the water dipole moments. Over the full range of salt concentrations, the orientations of water molecules within successive hydration shells are highly correlated, from layer to layer, through hydrogen bonding. Thus, the relatively large negative change in solvent entropy is attributed to the emergence of order of the ion shell structure upon base stacking, rather than the restructuring of the water dipole moments. These findings provide a more detailed picture in the context of the solvophobic bonding model, in which the enthalpic base stacking interaction is closely balanced by the decrease in entropy of the solvent environment. In contrast, this behavior is not observed for the singly-charged phosphate anion in isolation, suggesting that the presence of the bases in dApdA structure may be responsible for the disruption of the ion shell structure upon base un-stacking.

We also find that the trend in base stacking stability with increasing monovalent ion concentration for the dApdA dinucleotide is consistent with trends observed for the more complex duplex DNA (16,30). Although other factors, such as H-bonding and the cooperative stacking of multiple bases are known to play an important role in determining the stability of dsDNA structures, our results suggest that the restructuring of the ion shells about the central phosphate ion with increasing salt concentration, observed in dApdA, may also play a role in regulating the stability of larger DNA macromolecules.

DATA AVAILABILITY

Data discussed in the paper have been made available in Zenodo with reference number 3971255 (https://doi.org/10.5281/zenodo.3971255).

Source codes for the calculations of the CD spectra using the Extended Dipole Approximation and the Point Dipole Approximation have been made available in GitHub at https://github.com/GuenzaLab/dApdA.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors are grateful to Dr Pablo G. Romano for his contributions in early phases of this project. We are also grateful for many helpful discussions with the members of the Guenza, Marcus and von Hippel research groups.

FUNDING

National Science Foundation [CHE-1665466, CHE-1362500 to M.G.G.]; John Keana Fellowship (University of Oregon to E.R.B.); the computational work was performed on the supercomputer COMET at the San Diego Supercomputer Center, with the support of XSEDE allocation [TG-CHE100082 to M.G.G.] (The Extreme Science and Engineering Discovery Environment (XSEDE) (65) is supported by the National Science Foundation [ACI-1548562]); NSF [CHE-1608915 to A.H.M.]; NIH-NIGMS [GM-15792 to A.H.M. and P.H.v.H.]. Funding for open access charge: the University of Oregon.

Conflict of interest statement. None declared.

REFERENCES

1. von Hippel,P.H., Johnson,N.P. and Marcus,A.H. (2013) 50 Years of DNA 'breathing': reflections on old and new approaches. *Biopolymers*, **99**, 923–954.
2. Peyrard,M., Cuesta-López,S. and James,G. (2009) Nonlinear analysis of the dynamics of DNA breathing. *J. Biol. Phys.*, **35**, 73–89.
3. Alberts,B., Johnson,A.D., Lewis,J., Morgan,D., Raff,M., Roberts,K. and Walter,P. (2014) In: *Molecular Biology of the Cell*, 6th edn., Garland Science, NY.
4. Bosaeus,N., Reymer,A., Beke-Somfai,T., Brown,T., Takahashi,M., Wittung-Stafshede,P., Rocha,S. and Nordén,B. (2017) A stretched conformation of DNA with a biological role? *Q. Rev. Biophys.*, **50**, e11.
5. Feng,B., Sosa,R.P., Mårtensson,A.K.F., Jiang,K., Tong,A., Dorfman,K.D., Takahashi,M., Lincoln,P., Bustamante,C.J., Westerlund,F. *et al.* (2019) Hydrophobic catalysis and a potential biological role of DNA unstacking induced by environment effects. *Proc. Natl. Acad. Sci. U.S.A.*, **116**, 17169–17174.
6. Frank-Kamenetskii,M. (1987) How the double helix breathes, *Nature*, **328**, 17–18.
7. Williams,A.L., Cheong,C., Tinoco,I. and Clark,L.B. (1986) Vacuum ultraviolet circular dichroism as an indicator of helical handedness in nucleic acids. *Nucleic Acids Res.*, **14**, 6649–6659.
8. Cech,C.L., Hug,W. and Tinoco,I. (1976) Polynucleotide circular dichroism calculations: use of an all-order classical coupled oscillator polarizability theory. *Biopolymers*, **15**, 131–152.
9. Bayley,P.M., Nielsen,E.B. and Schellman,J.A. (1969) Rotatory properties of molecules containing two peptide groups: theory. *J. Phys. Chem.*, **73**, 228–243.
10. Lowe,M.J. and Schellman,J.A. (1972) Solvent effects on dinucleotide conformation. *J. Mol. Biol.*, **65**, 91–109.
11. Rizzo,V. and Schellman,J.A. (1984) Matrix-method calculation of linear and circular dichroism spectra of nucleic acids and polynucleotides. *Biopolymers*, **23**, 435–470.
12. Bush,C.A. and Tinoco,I. (1967) Calculation of the optical rotatory dispersion of dinucleoside phosphates. *J. Mol. Biol.*, **23**, 601–614.
13. Johnson,W.C. and Tinoco,I. (1969) Circular dichroism of polynucleotides: a general method applied to dimers. *Biopolymers*, **8**, 715–731.
14. DeVoe,H. and Tinoco,I. (1962) The stability of helical polynucleotides: base contributions. *J. Mol. Biol.*, **4**, 500–517.
15. Cantor,C.R., Warshaw,M.M. and Shapiro,H. (1970) Oligonucleotide interactions. 3. Circular dichroism studies of the conformation of deoxyoligonucleotides. *Biopolymers*, **9**, 1059–1077.
16. SantaLucia,J. (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl Acad. Sci. U.S.A.*, **95**, 1460–1465.
17. SantaLucia,J. and Hicks,D. (2004) The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct.*, **33**, 415–440.
18. Johnson,N.P. and Schleich,T. (1974) Circular dichroism studies of the conformational stability of dinucleoside phosphates and related compounds in aqueous neutral salt solutions. *Biochem*, **13**, 981–987.
19. von Hippel,P.H. and Wong,K.-Y. (1964) Neutral salts: the generality of their effects on the stability of macromolecular conformations. *Science*, **145**, 577–580.
20. Bowman,G.R., Pande,V.S. and Noé,F. (eds) (2014) In: *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation (Advances in Experimental Medicine and Biology)*. Vol. **797**, Springer, Dordrecht, Netherlands.
21. Chodera,J.D., Singhal,N., Pande,V.S., Dill,K.A. and Swope,W.C. (2007) Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics. *J. Chem. Phys.*, **126**, 155101.
22. Beyerle,E. and Guenza,M.G. (2019) Kinetics analysis of ubiquitin local fluctuations with Markov state modeling of the LE4PD normal modes. *J. Chem. Phys.*, **151**, 164119.
23. Rodger,A. and Nordén,B. (1997) In: *Circular Dichroism and Linear Dichroism, Oxford Chemistry Masters*, Oxford University Press, NY.
24. Ji,H., Johnson,N.P., von Hippel,P.H. and Marcus,A.H. (2019) Local DNA base conformations and ligand intercalation in DNA constructs containing optical. *Probes Biophys J*, **117**, 1101–1115.

Downloaded from https://academic.oup.com/nar/article/49/4/1872/6121468 by guest on 01 July 2021

25. Metzner,P., Schütte,C. and Vanden-Eijnden,E. (2006) Illustration of transition path theory on a collection of simple exemples. *J. Chem. Phys.*, **125**, 084110.
26. Metzner,P., Schütte,C. and Vanden-Eijnden,E. (2009) Transition path theory for Markov jump processes. *Multiscale Model. Simul.*, **7**, 1192–1219.
27. Hawk,A.T. and Makarov,D.E. (2011) Milestoning with transition memory. *J. Chem. Phys.*, **135**, 224109.
28. Meng,Y., Shukla,D., Pande,V.S. and Roux,B. (2016) Transition path theory analysis of C-Src kinase activation. *Proc. Natl Acad. Sci. U.S.A.*, **113**, 9193–9198.
29. Weinan,E. and Vanden-Eijnden,E. (2006) Towards a theory of transition paths. *J. Stat. Phys.*, **123**, 503–523.
30. Kitayner,M., Rozenberg,H., Rohs,R., Suad,O., Rabinovich,D., Honig,B. and Shakked,Z. (2010) Diversity in DNA recognition by P53 revealed by crystal structures with Hoogsteen base pairs. *Nat. Struct. Mol. Biol.*, **17**, 423–429.
31. Honig,B. and Rohs,R. (2011) Flipping Watson and Crick. *Nature*, **470**, 472–473.
32. Nikolova,E.N., Kim,E., Wise,A.A., O'Brien,P.J., Andricioaei,I. and Al-Hashimi,H.M. (2011) Transient Hoogsteen base pairs in canonical duplex DNA. *Nature*, **470**, 498–502.
33. Alvey,H.S., Gottardo,F.L., Nikolova,E.N. and Al-Hashimi,H.M. (2014) Widespread transient Hoogsteen base pairs in canonical duplex DNA with variable energetics. *Nat. Commun.*, **5**, 4786.
34. Zhou,H., Hintze,B.J., Kimsey,I.J., Sathyamoorthy,B., Yang,S., Richardson,J.S. and Al-Hashimi,H.M. (2015) New insights into Hoogsteen base pairs in DNA duplexes from a structure-based survey. *Nucleic. Acids. Res.*, **43**, 3420–3433.
35. Kypr,J., Kejinovská,I., Renciuk,D. and Vorlícková,M. (2009) Circular dichroism and conformational polymorphism of DNA. *Nucleic. Acids. Res.*, **37**, 1713–1725.
36. Rohs,R., West,S.M., Sosinsky,A., Liu,P., Mann,R.S. and Honig,B. (2009) The role of DNA shape in protein-DNA recognition. *Nature*, **461**, 1248–1253.
37. Coman,D. and Russu,I.M. (2005) A nuclear magnetic resonance investigation of the energetics of basepair opening pathways in DNA. *Biophys. J.*, **89**, 3285–3292.
38. Huguet,J.M., Bizarro,C.V., Forns,N., Smith,S.B., Bustamante,C. and Ritort,F. (2010) Single-molecule derivation of salt dependent base-pair free energies in DNA. *Proc. Natl Acad. Sci. U.S.A.*, **107**, 15431–15436.
39. Schildkraut,C. and Lifson,S. (1965) Dependence of the melting temperature of DNA on salt concentration. *Biopolymers*, **3**, 195–208.
40. 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. *Biochem*, **43**, 3537–3554.
41. Cantor,C.R. and Schimmel,P.R. (1980) In: *Biophysical Chemistry. Part I: The Conformation of Biological Macromolecules*. W. H. Freeman Ed.
42. Abraham,M.J., Murtola,T., Schulz,R., Pall,S., Smith,J.C., Hess,B. and Lindahl,E. (2015) GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, **1–2**, 19–25.
43. Duan,Y., Wu,C., Chowdhury,S., Lee,M.C., Xiong,G., Zhang,W., Yang,R., Cieplak,P., Luo,R., Lee,T. *et al.* (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.*, **24**, 1999–2012.
44. Jorgensen,W.L., Chandrasekhar,J., Madura,J., Impey,R. and Klein,M. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **79**, 926–935.
45. Bussi,G., Donadio,D. and Parrinello,M. (2007) Canonical sampling through velocity rescaling. *J. Chem. Phys.*, **126**, 14101.
46. Hess,B., Bekker,H., Berendsen,H.J.C. and Fraaije,J.G.E.M. (1997) LINCS: a linear constraint solver for molecular simulations. *J. Comp. Chem.*, **18**, 1463–1472.
47. Scherer,M.K., Trendelkamp-Schroer,B., Paul,F., Pérez-Hernández,G., Hoffmann,M., Plattner,N., Wehmeyer,C., Prinz,J.H. and Noé,F. (2015) 'PyEMMA 2: a software package for estimation, validation, and analysis of Markov models'. *J. Chem. Theory Comput.*, **11**, 5525–5542.
48. Arthur,D. and Vassilvitskii,S. (2007) In: *'k-means++: The Advantages of Careful Seeding' Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*. Philadelphia, PA, pp. 1027–1035.
49. Celebi,M.E., Kingravi,H.A. and Vela,P.A. (2013) A comparative study of efficient initialization methods for the K-means clustering algorithm. *Expert Syst. Appl.*, **40**, 200–210.
50. Prinz,J.-H., Wu,H., Sarich,M., Keller,B., Senne,M., Held,M., Chodera,J., Schütte,C. and Noé,F. (2011) Markov models of molecular kinetics: generation and validation. *J. Chem. Phys.*, **134**, 174105.
51. Reichl,L.E. (2016) In: *A Modern Course in Statistical Physics*. 4th edn, Wiley-VCH, Weinheim.
52. van Kampen,N.G. (2007) In: *Stochastic Processes in Physics and Chemistry*. 3rd edn, North Holland, Amsterdam, Boston.
53. Hamaguchi,K. and Geiduschek,E.P. (1962) The effect of electrolytes on the stability of the deoxyribonucleate helix. *J. Am. Chem. Soc.*, **84**, 1329–1338.
54. Manning,G.S. (1969) Limiting laws and counterion condensation in polyelectrolyte solutions. I. Colligative properties. *J. Chem. Phys.*, **51**, 924–933.
55. Manning,G.S. (1969) Limiting laws and counterion condensation in polyelectrolyte solutions. II. self-diffusion of the small ions. *J. Chem. Phys.*, **51**, 934–938.
56. Lo Nostro,P. and Ninham,B.W. (2012) Hofmeister phenomena: an update on ion specificity in biology. *Chem. Rev.*, **112**, 2286–2322.
57. Bai,Y., Greenfeld,M., Travers,K.J., Chu,V.B., Lipfert,J., Doniach,S. and Herschlag,D. (2007) Quantitative and comprehensive decomposition of the ion atmosphere around nucleic acids. *J. Am. Chem. Soc.*, **129**, 14981–14988.
58. Lipfert,J., Doniach,S., Das,R. and Herschlag,D. (2014) Understanding nucleic acid–ion interactions. *Annu. Rev. Biochem.*, **83**, 813–841.
59. Romano,P.G. and Guenza,M.G. (2017) GRadient adaptive decomposition (GRAD) method: optimized refinement along macrostate borders in markov state models. *J. Chem. Inf. Model.*, **57**, 2729–2740.
60. Röblitz,S. and Weber,M. (2013) Fuzzy spectral clustering by PCCA+: application to Markov state models and data classification. *Adv. Data Anal. Classi*, **7**, 147–179.
61. Privalov,P.L. (2020) Physical basis of the DNA double helix. *JBSB*, **8**, 1–7.
62. Baldwin,R.L. (2014) Dynamic hydration shell restores Kauzmann's 1959 explanation of how the hydrophobic factor drives protein folding. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 13052–13056.
63. Hoel,P.G., Port,S.C. and Stone,C.J. (1986) In: *Introduction to Stochastic Processes*. Waveland Press.
64. Bouvier,B., Gustavsson,T., Markovitsi,D. and Milli,P. (2002) Dipolar coupling between electronic transitions of the DNA bases and its relevance to exciton states in double helices. *Chem. Phys.*, **275**, 75–92.
65. Towns,J., Cockerill,T., Dahan,M., Foster,I., Gaither,K., Grimshaw,A., Hazlewood,V., Lathrop,S., Lifka,D., Peterson,G.D. *et al.* (2014) XSEDE: Accelerating Scientific Discovery. *Comput. Sci. Eng.*, **16**, 62–74.

Downloaded from https://academic.oup.com/nar/article/49/4/1872/6121468 by guest on 01 July 2021