PCCP

Physical Chemistry Chemical Physics

Accepted Manuscript

Published on 13 November 2026
Licensed under CC-BY-NC 4.0

Check for updates



This is an Accepted Manuscript, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this Accepted Manuscript with the edited and formatted Advance Article as soon as it is available.

You can find more information about Accepted Manuscripts in the Information for Authors.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard Terms & Conditions and the Ethical guidelines still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this Accepted Manuscript or any consequences arising from the use of any information it contains.

ROYAL SOCIETY OF CHEMISTRY

rsc.li/pccp

PCCP

ARTICLE TYPE

Cite this: DOI: 00.0000/xxxxxxxxxx

Modeling IR Spectral Signatures of Solvated Single-Stranded DNA: Linking C=O Stretching Vibrations to Conformational Dynamics†

Cedrix J. Dongmo Foumthuim,[a] Giovanni Parolin,[a] Stefano Corni,[a,b] and Laura Zanetti-Polzi*[b]

Received Date
Accepted Date

DOI: 00.0000/xxxxxxxxxx

The carbonyl stretching mode is a widely used signal in the infrared spectroscopy of biomolecules, being highly sensitive to changes in their structural organization, such as secondary structure in peptides and proteins or distinct structural motifs in nucleic acids. However, direct mapping between IR spectral bands and specific structural features is not straightforward, particularly for systems with a rugged conformational landscape. Here, we present a computational study of the infrared spectra of C=O stretching modes in DNA bases using a hybrid quantum–classical approach based on the perturbed matrix method. This methodology combines classical molecular dynamics simulations with quantum mechanical calculations on the isolated chromophore, enabling the modeling of vibrational signals across large ensembles of conformations. Recently, we applied the same approach to model the C=O stretching band in thymine-based systems [Parolin et al. PCCP 27.37(2025): 19797-19808]. Here, we expand on that work by investigating guanine- and cytosine-derived systems, thus completing the set of DNA bases containing carbonyl groups and enabling the calculation of C=O stretching in any DNA system. We first benchmark the method on 9-methylguanine and 1-methylcytosine, then extend it to guanine and cytosine 5'-monophosphates, single-stranded deca-homonucleotides, and a seven-base single-stranded DNA oligomer containing all nucleobases. The calculated spectra well reproduce the experimental observations, including frequency shifts and intensity variations arising from stacking and environmental fluctuations, and highlight the role of excitonic effects in shaping spectral features. These results show that the methodology provides insights into how local interactions and conformational dynamics determine the IR spectral signatures of DNA bases.

1 Introduction

Establishing a quantitative connection between experimentally measured spectroscopic signals and their theoretical counterparts is a non-trivial task. In principle, computational approaches have the potential to access detailed structure–spectra relationships that are often difficult to disentangle experimentally. Nonetheless, the accurate reproduction of experimental spectroscopic observables requires accounting for multiple physical processes. Specifically, spectroscopic signals arise from the interplay between excitation events, their possible mutual couplings, and the dynamic fluctuations of the surrounding environment. Moreover, spectroscopic measurements typically probe ensemble-averaged responses over a broad distribution of conformational states, such that the observed spectral features reflect both static structural heterogeneity and time-dependent fluctuations occurring across multiple length and time scales. As a consequence, a reliable comparison between calculated and experimental spectra requires modeling both the intrinsic response of individual chromophores to excitation and the effects induced by environmental dynamics and inter-chromophore interactions.

Among spectroscopic techniques such as nuclear magnetic resonance, Forster resonance energy transfer, and circular dichroism, infrared (IR) spectroscopy is a widely used and powerful method that is particularly sensitive to local structural motifs and to their interactions with the surrounding environment. Owing to its applicability under physiological conditions, IR spectroscopy is extensively employed in the study of biomolecules to probe their conformational landscapes. However, while the high sensitivity of IR spectroscopy provides access to a wealth of information on structure–function relationships, much of this information is often hidden within broad, overlapping vibrational

[a] Department of Chemical Sciences, University of Padova, Padova 35131, Italy
[b] Institute of Nanoscience - CNR, S3, via Campi 213/A, 41125 Modena, Italy. Email: laura.zanettipolzi@cnr.it

† Supplementary Information available: additional figures; output/input files for quantum mechanical calculations and molecular dynamics simulations. See DOI: 10.1039/cXCP00000x/

Physical Chemistry Chemical Physics Accepted Manuscript

bands that are difficult to disentangle experimentally. The computational modeling of molecular vibrations in biomolecules has significantly contributed to the interpretation of experimental spectra. Approaches such as the Hessian Matrix Reconstruction method were used to investigate the sensitivity of IR spectra to various fundamental factors such as delocalization of mode couplings, deuterium isotopic exchange effects, and electrostatic interactions among neighboring chromophores both in the context of (poly)peptides [1–3] and in that of nucleic acids. [4–7] Computational approaches based on frequency maps, i.e. semiempirical models developed to reconstruct vibrational spectra for specific IR modes on the basis of *ab initio* calculations and experimental data, are also widely used for the modeling of signals arising from polypeptides and nucleic acids. [8–11] More recently, approaches embedding machine learning paradigms have shown promise in reconciling IR spectra modeling and experimental data. [12,13]

Aiming at providing a physically ground computational approach with a specific focus on the interplay between the molecular vibrations and the complex dynamics of the environment, we used in the past years the MD-PMM approach [14] to model IR signals from biomolecules. The MD-PMM is a tandem quantum-classical methodology based on the approximation of dividing the system into a region that is treated at the quantum mechanical (QM) level (in the case of IR spectra calculations, the vibrating chromophore) and a complementary region (the environment) that is treated at a classical level and perturbs electrostatically the quantum region. The salient feature of the approach is that it allows us to calculate the IR absorption signal arising from a large number of chromophore+environment conformations (i.e. all snapshots obtained from a classical MD simulation). The approach does not require the use of adjustable parameters to model the environment-dependent spectral changes. As a matter of fact, it only applies a rigid shift accounting for inaccuracies of the QM calculations, i.e, compensating the discrepancy between the calculated gas phase frequency and the corresponding experimental one in Ar/N_2 matrices and the non-optimal description of the eigenvector associated to the mode of interest (*vide infra*). This approach has already been used satisfactorily to compute the IR spectra in polypeptides, both in the amide I region [15–18] and for specific side chains bands. [19,20] Very recently, we extended it to account for the vibrational dynamics of carbonyl in-plane stretching modes in DNA, and specifically in systems derived from thymine. [21] The C=O stretching mode, found in all nucleobases but adenine, is highly responsive to fluctuations in the structural motifs of nucleic acids and has been used to connect different conformational patterns of DNA to frequency shifts and/or intensity changes. [22,23]

In this paper, we use the MD-PMM to compute the IR signals of the C=O stretching bands in guanine and cytosine bases, thus completing the calculation of C=O bands in DNA bases we began with thymine in our previous work. [21] Specifically, we address the calculations of the bands resulting from the C=O vibrating mode in 9-methylguanine (9MG) and 1-methylcytosine (1MC) (see Figure 1); guanidine 5'-monophosphate (GMP) and cytosine 5'-monophosphate (CMP); and 10-bases single-stranded oligonucleotides of guanine (dG10, 5'-GGGGGGGGGG-3') and cytosine (dC10, 5'-CCCCCCCCCC-3'), both in water and in D_2O. Moreover, we also consider a single-stranded DNA oligomer, 7merDNA (5'-AGGTTCA-3') embedding all nucleobases, including thymine, and whose experimental IR spectrum is known. [24] Overall, the calculated spectra compare well with the corresponding experimental ones, thus reinforcing the ability of the proposed approach to adequately reproduce the subtle experimentally observed variations in the spectral signals triggered by the conformational diversity around the considered chromophore, including solvation, stacking, and hydrogen bonding interactions among the bases.

2 Materials and methods

2.1 The MD-PMM approach

The MD-PMM approach has been widely applied to the calculation of IR signals, mainly in the context of peptides and proteins, and more recently to reconstruct the two C=O stretching modes of the thymine DNA base. Its theoretical ground, extensively discussed in previous papers, [14,25,26] is based on treating the vibrating chromophore (the quantum center, QC) at the quantum mechanical (QM) level and considering the electrostatic effect of the environment on the electronic states of the QC. Practically, the unperturbed properties of the QC alone are initially computed in the gas phase. Subsequently, the effect of the perturbing environment is accounted for by applying a perturbation term $\widehat{V}$ for each environment configuration as obtained from a classical MD simulation. Therefore, the electronic Hamiltonian operator $\widehat{H}$ of the QC embedded in the perturbing environment of configuration k is described as:

$$\widehat{H}_k = \widehat{H}_0 + \widehat{V}_k \tag{1}$$

where $\widehat{H}_0$ is the QC unperturbed electronic Hamiltonian and the components $V_{i,j}$ of the perturbation operator between the electronic states i and j are given by:

$$V_{i,j} = \delta_{i,j} \sum_N \mathscr{V}(\mathbf{R}_N) q_{N,i} - (\boldsymbol{E}(\mathbf{r}_0) \cdot \boldsymbol{\mu}_{i,j})(1 - \delta_{i,j}) \tag{2}$$

with N running over all QC atoms. $\boldsymbol{R}_N$ is the nucleus position of the Nth atom of the QC, $q_{N,i}$ is its atomic charge in the state i, $\mathscr{V}$ is the electrostatic potential exerted by the perturbing environment, $\boldsymbol{E}$ is the perturbing electric field ($\boldsymbol{E} = -\partial\mathscr{V}/\partial\mathbf{r}$) evaluated at the center of mass $\mathbf{r}_0$ of the QC, $\boldsymbol{\mu}_{i,j}$ is the QC transition dipole moment and $\delta_{i,j}$ is the Kronecker delta.

The computational pipeline can be summarized as follows:

1. Quantum mechanical calculations are performed on the optimized geometries of the QCs, enabling us to obtain the mass-weighted Hessian eigenvectors. In this work, 9-methylguanine (9MG) and 1-methylcytocine (1MC) are used as QCs (see section 2.2). Then, the modes of interest are selected (in the present case, the C=O stretching modes in 9MG and 1MC).

2. A number of QC configurations are extracted along the eigenvector corresponding to the mode of interest. For each of these configurations, QM calculations are performed, providing the unperturbed properties of the QC. (see section

Physical Chemistry Chemical Physics Accepted Manuscript

2.2).

3. A classical MD simulation of the whole system (QC + environment) is performed. In this work, we perform MDs of the following systems in water: 9MG, GMP, dG10 (for guanine-related systems); 1MC, CMP and dC10 (for cytosine-related systems); and 7merDNA oligomer (see section 2.3).

4. The perturbed ground state energy of the QC is calculated for each snapshot of the MD simulation. This is done by constructing and diagonalizing the perturbed Hamiltonian matrix (Eq. 1). It is noteworthy that this is done for all the QC configurations generated at point 2. Moreover, it should be emphasized that for calculating the spectrum of dG10, dC10, and 7merDNA, this procedure is looped over every single base of the oligomer. In doing so, the first base of the DNA is considered QC while the remaining bases along with the solvent and counterions represent the perturbing environment, leading to the IR signal of the first base, then the second base is considered and so on.

5. The perturbed energy curve deriving from each MD frame (and, for the oligomers, for each base) is modeled by a Morse potential that results in the perturbed vibrational frequency for each MD frame. This fitting procedure also allows for estimation of the anharmonic contribution to the perturbed frequency. A histogram of these frequencies, weighted by the corresponding transition dipole moments, provides the IR vibrational spectrum of the mode of interest for the target chromophore embedded in its perturbing environment.

To fully describe the oligomers considered herein (dG10, dC10, and 7merDNA), the interactions of different vibrational centers (i.e., modes coupling) have to be included in the calculations. This is done by constructing the excitonic coupling matrix at each MD frame, taking into account the coupling among modes at the transition-dipole coupling (TDC) level of approximation, i.e., by expanding the chromophore-chromophore interaction operator up to the dipolar terms. In our specific case, coupling effects are taken into account among the C=O stretching modes of the ten bases of the dG10 and dC10 oligomers, respectively (intra-mode coupling, i.e. coupling among the same vibrational mode in different chromophores). In the case of the 7merDNA oligomer, mode coupling must also include inter-mode couplings (i.e., coupling between different chromophores and different vibrational modes).[21]

The excitonic coupling Hamiltonian matrix is defined as follows:[26]

$$\widetilde{H} = \tilde{I}\mathscr{V}_{vb}^{0} + \Delta\widetilde{H} \tag{3}$$

where $\mathscr{V}_{vb}^{0}$ is the vibrational ground state energy of the interacting chromophores and $\Delta\tilde{H}$ is the excitation matrix, whose diagonal elements are:

$$\left[\Delta\widetilde{H}\right]_{k,k} = h\nu_k \tag{4}$$

and whose non-zero off-diagonal elements are given by the chromophores interaction operator (at the dipolar approximation):

$$\hat{V}_{k,k'(k\neq k')} = \frac{\hat{\boldsymbol{\mu}}_k \cdot \hat{\boldsymbol{\mu}}_{k'}}{R_{k,k'}^3} - 3\frac{(\hat{\boldsymbol{\mu}}_k \cdot \mathbf{R}_{k,k'})(\hat{\boldsymbol{\mu}}_{k'} \cdot \mathbf{R}_{k,k'})}{R_{k,k'}^5} \tag{5}$$

In the above equations, ν_k is the kth chromophore excitation frequency, $\hat{\boldsymbol{\mu}}_k$ the kth chromophore dipole operator and $\mathbf{R}_{k,k'}$ is the k' to k chromophore displacement vector. By diagonalizing the above matrix and using the transition dipole for the $0 \rightarrow i$ excitonic transition ($\boldsymbol{\mu}_{0,i}$) obtained via the excitonic eigenvectors, we reconstruct the spectral signal of the excitonic system by summing the absorbance due to each $0 \rightarrow i$ transition:

$$\varepsilon(\nu) = \sum_i \frac{|\boldsymbol{\mu}_{0,i}|^2 \rho_i(\nu) h\nu}{6\varepsilon_0 c\hbar^2} \tag{6}$$

with ρ_i the probability density in ν frequency space for the ith excitation and ε_0 the vacuum dielectric constant.

2.2 QM calculations

The QCs used in the framework of the MD-PMM approach in this work are 9-methylguanine (9MG), 1-methylcytosine (1MC), and their deuterated analogs. Specifically, QM calculations are carried out on isolated 9MG and 9MG-D (i.e., 9MG deuterated at N_1 and N_2, see Figure 1B); 1MC and 1MC-D (i.e., 1MC deuterated at N_4, see Figure 1C), at the density functional theory (DFT)[28,29] level with the 6-31+G(d) basis set[30,31] in conjunction with the B3LYP functional[32] and the D3BJ dispersion correction,[33] i.e., at the same level of theory previously employed for the calculation of infrared spectra of polypeptides and thymine-related DNA base with the MD-PMM approach.[19–21,25]

The mass-weighted Hessian matrix is calculated on the optimized geometry of 9MG(-D) and 1MC(-D) at the B3LYP+D3BJ/6-31+G(d) level of theory and subsequently diagonalized to obtain the unperturbed eigenvectors and related eigenvalues. The eigenvectors corresponding to the C=O stretching (see Figures 1B and C and S1 in the Electronic Supplementary Information, ESI) are then used to generate 26 points (configurations) along each C=O mode using a step of 0.02 a.u., and spanning an energy range of $\approx$ 50 kJ/mol. For each point, calculations on the ground state and the lowest three excited states (at the time-dependent DFT level[34]) are performed using the same functional and basis set, leading to the unperturbed properties, used as input for the MD-PMM calculations. All QM calculations are performed using the Gaussian16 package.[35]

2.3 MD simulations

Atomistic MD simulations for the model systems MX, XMP and dX10 (X=G, C) and the 7merDNA oligomer are performed with the CHARMM36 force field,[40–42] combined with the modified version of the TIP3P water model.[43,44] The force field parameters for MX models (9MG and 1MC) are obtained from the CHARMM general force field (CGenFF).[45] The starting coordinates of MX and XMP (GMP and CMP) structures are prepared using the Avogadro tool (version 1.2.0).[46] The force field parameters for XMP are manually tuned from their corresponding de-

Physical Chemistry Chemical Physics Accepted Manuscript

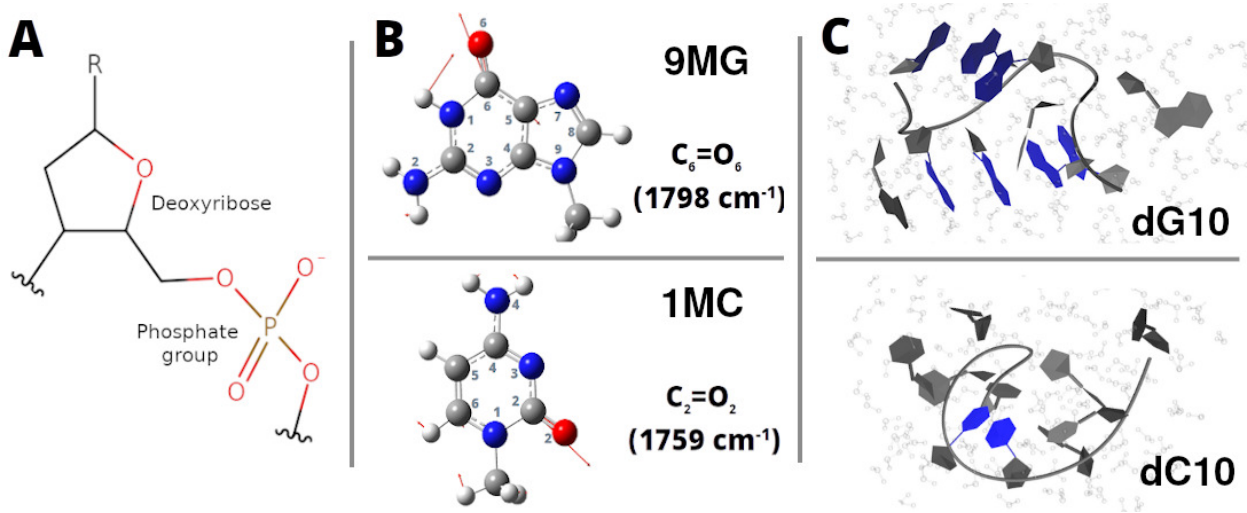


Fig. 1 A: Schematic illustration of DNA structure with evidence of the phosphodiester link bridging the nucleobases R (R = A, C, G or T). B: When the sugar-phosphate backbone is replaced by a methyl group at N_9 in guanine, it results to 9-methylguanine (9MG); when it is replaced at N_1 in cytosine, the resulting structure is 1-methylcytosine (1MC). The carbonyl C=O stretching mode eigenvectors are reported as red arrows and the corresponding gas-phase frequencies are reported on the top. A comparable picture for the deuterated bases 9MG-D and 1MC-D, i.e. when protons at N_1 and N_2 in 9MG and at N_4 in 1MC are substituted by deuterium atoms, is reported in Fig. S1 in the ESI. Atoms are labeled following the IUPAC-IUB numbering scheme.[27] C: Representative snapshots of the oligomers dG10 and dC10. Stacked bases are shown in blue and the first solvation shell is depicted by the grayscale water molecules.

Table 1 Experimental and calculated frequencies in cm^{-1} for 9MG, GMP, 1MC and CMP.

	Calc. gas phase	Exp. Ar/N_2	Rigid shift	Calc. in solvent	Calc. in solvent shifted	Exp. in solvent	Calc. shift upon solvation	Exp. shift upon solvation
9MG	1798	1741[a]	57	1765	1708	-	33	-
9MG-D	1791	1736[a]	55	1758	1703	-	33	-
GMP*	1791	1736[a]	55	1742	1687	1665[d]	49	71
1MC	1759	1718[b,c]	41	1690	1649	-	69	-
1MC-D	1758	-	41	1690	1649	-	68	-
CMP*	1758	-	41	1669	1628	1651[d]	89	66

* Gas phase calculations are performed on 9MG-D and 1MC-D for the calculation of the spectra of GMP and CMP, respectively. [a]From Sheina *et al.* [36]. [b,c] Value averaged over reference ones from [b] Szczesniak *et al.* [37] and [c] Szczesniak *et al.* [38]. [d] From Peng *et al.* [39]

fault CHARMM36 parameters. Specifically, atoms O2' and H2' are deleted and replaced by a single hydrogen atom, and their corresponding charges are summed up to the remaining atoms, thereby preserving the overall initial charge of −2. The initial coordinates of the dX10 (dG10 and dC10) and 7merDNA oligomers are obtained by generating a double helical structure of B-DNA and then removing the corresponding complementary chains. Two, nine and six Na^+ counterions are added to the simulation box of XMP, dX10 (X=G, C) and 7merDNA, respectively, to neutralize the negative charges of the backbone, while the C_3' and C_5' ends in dX10 and 7merDNA are capped with an OH group. For 9MG and dG10, simulations are also performed using the TIP4P/2005 water model to assess possible effects of the water model on the spectral features. The calculated spectra, reported in Figure **??** in the ESI, show that, for 9MG in water, the spectrum obtained with the TIP4P/2005 model is slightly broader and blue-shifted (from 1708 cm^{-1} to 1717 cm^{-1}). For dG10, the spectra obtained with the two water models are nearly superimposable.

Simulations are performed in the isothermal-isobaric ensemble (NPT), using the Bussi, Donadio, Parrinello thermostat[47] and the Parrinello-Rahman barostat:[48] the reference temperature and pressure are set at 298.15 k and 1 bar, respectively. After energy minimization, thermalization (200 ps) and equilibration under NPT conditions (500 ps), trajectories are propagated for 50 ns (MX), 100 ns (XMP), and 200 ns (dX10 and 7merDNA) with a time step of 2 fs. Coordinates are saved every 2 ps for the entire system, providing a statistically meaningful sampling of solute and solvent configurations. Following common prescriptions for the selected force field, all bonds involving H atoms are constrained using the LINCS algorithm.[49] Moreover, a 1.2 nm cutoff is chosen for both short-range electrostatic and van der Waals interactions. In the latter case, a force-switching function is applied between 1.0 nm and the cutoff, and no long-range correction is applied. The long-range electrostatic is taken into account us-

Physical Chemistry Chemical Physics Accepted Manuscript

ing the particle mesh Ewald scheme (PME).[50,51] The size of the simulation boxes allows for the inclusion of a sufficient number of solvent molecules to prevent any unphysical interactions between periodic replicas in adjacent cells (the minimum distance between replicas is at least 1.4 nm during the simulations) and to provide a converged electrostatic perturbation in the quantum center (QC). All MD simulations and preliminary analysis of the trajectories are performed with the Gromacs suite of packages[52] and the MDAnalysis library.[53]

3 Results and discussion

The structure of DNA consists of repeating nucleotide units, each comprising one of the four nucleobases - adenine (A), thymine (T), cytosine (C), or guanine (G) - attached to a sugar-phosphate backbone. The IR band corresponding to the C=O stretching mode in DNA arises from vibrations of carbonyl groups in thymine, guanine and cytosine, as adenine lacks such groups. The calculation of the two C=O stretching bands arising from thymine using the MD-PMM approach has previously been reported.[21] Here, the same method is applied to compute the bands associated with the carbonyl groups of guanine and cytosine, thereby completing the preliminary calculations required to enable the modeling of the C=O stretching signal in any DNA structure.

Following the strategy used for the two thymine bands in our previous work, we first design model systems that embed the methylated bases of guanine, 9-methylguanine (9MG), and cytosine, 1-methylcytosine (1MC) (Figure 1), to be used as quantum centers as described in section 2. In the frequency region 1500-1800 cm^{-1}, guanine and cytosine exhibit signals assigned to the in-plane stretching modes of carbonyl (C=O) and ring (C=C and C=N) vibrations.[23,54] Here, we focus on the first of these modes. Using 9MG and 1MC, we compute, for each base and its corresponding deuterium-substituted analog (see Figures 1B and C and Figure S1 in the ESI) the gas-phase normal-mode vibrations of carbonyl stretching at the harmonic approximation. We obtain 1798 cm^{-1}(1791 cm^{-1}) and 1759 cm^{-1}(1758 cm^{-1}) for 9MG (9MG-D) and 1MC (1MC-D), respectively. The experimental counterparts to our calculations on 9MG, 9MG-D, and 1MC are isolated low temperature data in argon and nitrogen matrices, which are 1741 cm^{-1},[36] 1736 cm^{-1},[36] and 1718 cm^{-1},[37,38] respectively (see Table 1). The calculated and Ar/N_2 matrix measurements differ by $\approx$ 50 cm^{-1}, which is inherent to this class of calculations and pertains both to the intrinsic inaccuracies of the QM calculations and to the use of the harmonic approximation.[19,21,26] In fact, the experimental C=O stretching frequencies of isolated 9MG, 9MG-D, and 1MC in Ar/N_2 matrices also embed anharmonic effects. By fitting the unperturbed energy curve to a Morse potential, we estimate that the anharmonic contribution to the unperturbed frequency is 10 cm^{-1}, 11 cm^{-1}, and 12 cm^{-1} for the C=O stretching modes of 9MG, 9MG-D and 1MC, respectively. Moreover, by fitting the perturbed energy curve of each MD frame to a Morse potential, the anharmonic contribution to the spectra calculated in water (see below) can also be obtained. This calculation shows that for the C=O stretching mode in all three compounds, the anharmonic contribution to the perturbed frequencies remains steadily constant along the frequency trajectory and only slightly unchanged compared to that in the gas-phase (see ESI, Figure S3). Therefore, the previously observed discrepancy between the calculated gas phase frequency and the corresponding experimental one in Ar/N_2 matrices is compensated for by applying a rigid frequency shift that accounts for both the anharmonicity and inaccuracies of the QM calculations. Consequently, each perturbed frequency obtained at the harmonic level is downshifted by 57 cm^{-1}, 55 cm^{-1}, and 41 cm^{-1} for the C=O stretching mode of 9MG, 9MG-D, and 1MC, as reported in Table 1. Since we could not find any experimental frequency for isolated 1MC-D, we apply to its calculated gas phase frequency the same shift applied to 1MC (Table 1) after verifying that the anharmonic contribution is equal to that obtained for 1MC and constant along the frequency trajectory (Figure S3 in the ESI).

In line with the MD-PMM computational workflow (see Section 2.1), we then include the electrostatic perturbation of the environment on the QC C=O stretching frequency, allowing us to obtain the perturbed vibrational spectra of the C=O stretching modes of 9MG (9MG-D) and 1MC (1MC-D) in water reported in Figure 2A and B, respectively. Since no direct experimental counterparts could be found for either the 9MG or the 1MC model systems (or deuterated analogs) we also calculate the C=O stretching bands in guanidine and cytosine 5'-monophosphate (GMP and CMP, i.e. the base bound to D-ribose-5-phosphate) for which the corresponding experimental spectra are available.[39] Since experimental spectra are acquired in D_2O, as commonly done to avoid the overlap between C=O stretching bands and the water bending mode ($\approx$ 1640 cm^{-1}), the computed spectra are obtained using as QCs the deuterated bases 9MG-D and 1MC-D together with the MD simulation of the corresponding base in water. As a matter of fact, explicit inclusion of heavy isotopic water D_2O in MD simulations was shown to induce limited effects on the equilibrium properties of biomolecular systems[55–57] and consequently on the calculated spectra. The experimental and calculated peak frequencies for the model systems XMP (X=G or C) in D_2O are summarized in Table 1 and the spectra are plotted in Figure 2 along with the corresponding experimental spectra.[39]

Comparison between the experimental and computed spectra shows that the calculated band of GMP is upshifted by 22 cm^{-1} with respect to the experimental band. The calculated band of CMP is instead downshifted by 23 cm^{-1} with respect to the experimental band. The difference between the experimental and calculated frequencies is similar in magnitude to that obtained in our previous work for the C_4=O stretching band in thymine.[21] In that work, a thorough analysis of the origin of the discrepancy was conducted, showing that it arises from a non-optimal description of the eigenvector associated to the mode of interest in the QM calculations. Given that our main aim is to reproduce the spectral changes due to the surrounding environment (e.g. structural arrangements, hydrogen bonding patterns, and interaction with water) and in line with our previous work, we accounted for the difference observed for GMP and CMP by applying an additional rigid shift of -22 cm^{-1} and +23 cm^{-1}, respectively, to all calculated signals arising from C=O stretching modes in G and C bases.

Then, we test the actual capability of the calculated signals to

Physical Chemistry Chemical Physics Accepted Manuscript

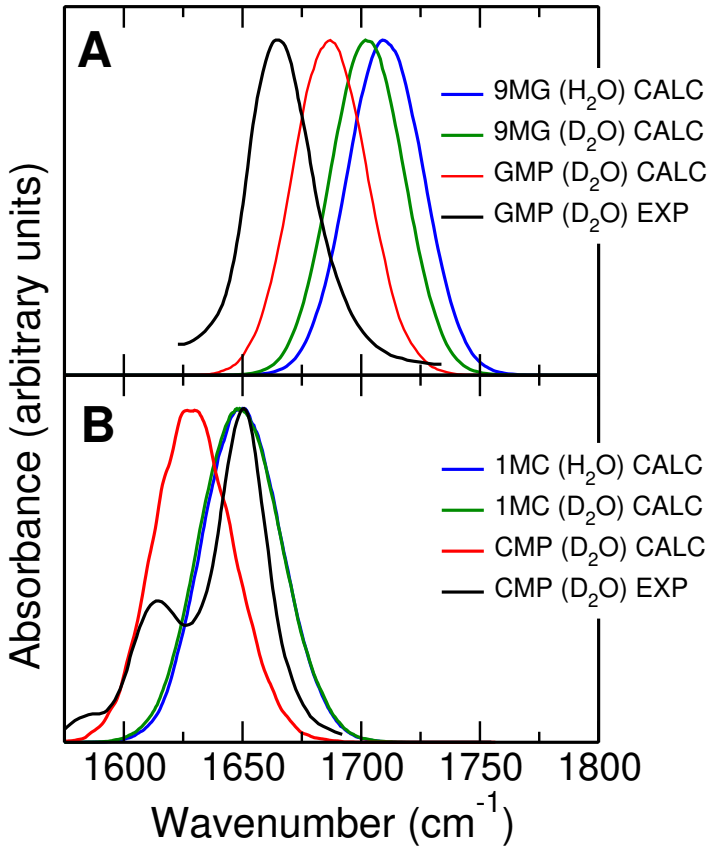


Fig. 2 Experimental and calculated C=O stretching bands for 9MG/GMP (A) and 1MC/CMP (B). Blue: calculated IR signal for 9MG and 1MC in water. Green: calculated IR signal for 9MG and 1MC in D_2O. Red: calculated IR signal for GMP and CMP in D_2O. Black: Experimental spectrum of GMP and CMP in D_2O.[39] Note that the low frequency band present in the experimental spectrum of CMP in B arises from the C=C and C=N stretching modes not modeled in our calculations.[39] The IR signals are normalized at the intensities of the highest peak.

reproduce the spectral changes due to environmental differences by addressing the calculation of the bands arising from C=O stretching in two single stranded decahomonucleotides dX10 (X=G or C), made up of 10 bases of either guanine (dG10: 5'-GGGGGGGGGG-3) or cytosine (dC10: 5'-CCCCCCCCCC-3). In line with the MD-PMM protocol described in Section 2.1, the C=O vibrational modes of dX10 in D_2O are obtained by using QM calculations on MX-D systems (9MG-D and 1MC-D), followed by the computation of the perturbed frequencies for each of the 10 bases of the oligomer along the MD trajectory of dX10. It should be noted that for single-base systems (i.e., MX and XMP) no modes coupling effects have to be included. Therefore, the corresponding bands are the histograms of the perturbed frequencies, weighted by the corresponding transition dipoles. In contrast, the final spectra of the dX10 oligomers also include mode coupling effects as described in Section 2.1. For the calculation of the dX10 spectra, only the last 140 ns of the MD simulations are used. In fact, the MD simulations start from extended, thus unrealistic conformations, which require tens of nanoseconds to relax (see the discussion below and Figures S4 and S5 in the ESI). The calculated and corresponding experimental spectra are reported in Figure 3 (XMP systems in black and dX10 oligomers in red, with X=G or C). The figure shows that the calculations reproduce well the differences observed experimentally when moving from single-base systems (XMP) to realistic DNA oligomers. The peak of the calculated C=O stretching band for dG10 is at 1674 cm^{-1} i.e., upshifted by 9 cm^{-1} with respect to GMP and well reproducing the experimentally observed shift (1665 cm^{-1} in GMP *vs* 1675 cm^{-1} in dG10). Also in line with the experimental spectra, no frequency shift is observed between the calculated spectra of CMP and dC10. It should also be noted that the experimental differences between the GMP/dG10 spectral lineshapes are reproduced in the calculated spectra only if the excitonic contribution is included in the calculation (see Figure S6 in the ESI). This observation suggests that the relative orientation of the bases in the oligomers plays a relevant role in the experimentally observed differences. This motivated further investigations of the interactions among the bases to correlate the spectral signals with specific structural patterns.

We analyze the stacking interactions among the bases in the dX10 oligomers. This was done by computing the center of mass (COM) distance d between two bases ring planes and the angle θ between their corresponding normal vectors. Following common practices,[58,59] two bases in the oligomer along the MD trajectory were considered stacked if $d \leq 4$ Å and $\theta \leq 35°$. The average stacking fraction per-residue for dG10 and dC10 is reported in Figure 4 (representative snapshots are shown in Figure 1C) showing that dG10 has a larger propensity to stacking interactions with respect to dC10 (the calculated average stacking fraction is approximately 0.43 ± 0.17 for dG10 and 0.21 ± 0.17 in dC10). This difference is consistent with what is generally expected in purine and pyrimidine systems, with pyrimidine bases showing a lower propensity for stacking interactions.[60] In addition, the plots show a heterogeneous pattern of the stacking fraction in dG10 while a nearly uniform trend among the bases is observed in dC10. This suggests a more complex and dynamic interaction scenario among the bases in dG10 than in dC10. In the case of dG10, the time dependent profiles of the structural order parameters shown in Figure S4 in the ESI show that, after ≈ 60 ns, the single-stranded oligomer reaches a reasonably equilibrated region of its conformational landscape (Figure S4A). This structural equilibration is associated with a coil-to-globule conformational transition, evidenced by the progressive decrease in the radius of gyration (Figure S4B), concomitantly with a steady increase in the number of intramolecular hydrogen bonds between the bases (Figure S4C) and in the stacking interaction among the different bases of the oligomer (Figure S4D). In dC10, the same structural parameters do not show any clear trend along the MD simulation (see Figure S5 in the ESI) with no sign of conformational transition or collapse, in line with the relatively small average stacking fraction observed (0.21 ± 0.17). However, for consistency across the analysis, the initial 60 ns of the MD trajectories were excluded from the spectral calculations for both dG10 and dC10.

To better understand the effect of stacking interactions on the spectral lineshape, we cluster the conformational ensemble of dG10 in sub-spaces using the average stacking fraction as a met-

Physical Chemistry Chemical Physics Accepted Manuscript

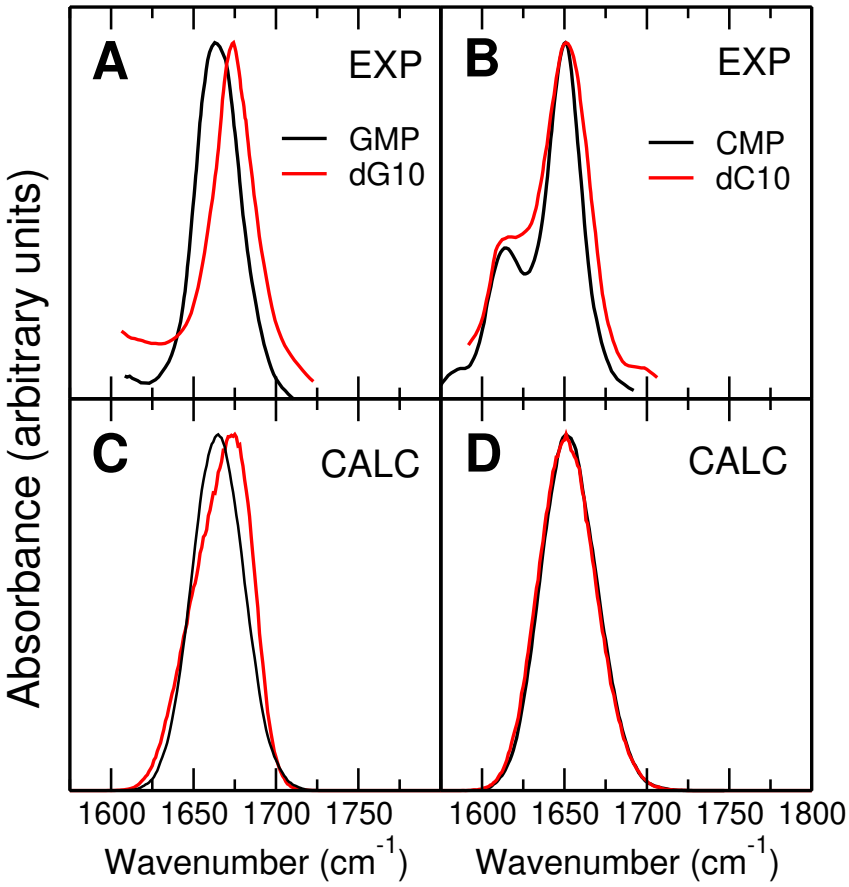


Fig. 3 Experimental (upper panels) and calculated (lower panels) spectra of XMP (black) and dX10 (red) model systems, where X stands either to guanine G or cytosine C. A/C and B/D panels refer to guanine and cytosine bases, respectively. Note that the low frequency band/shoulder present in the experimental spectrum of cytosine-related systems arises from the C=C and C=N stretching modes not modeled in our calculations. The IR signals are normalized at the intensities of the highest peak. Note that the calculated bands include the additional rigid shift discussed in the text.

ric. Specifically, we select trajectory intervals featuring a below-average (≈ 0.2) and above-average (≈ 0.6) stacking fraction, and compare the spectra obtained using these sub-populations with that obtained from the full trajectory (≈ 0.4 stacking fraction) and that obtained for GMP (as null stacking fraction reference). The corresponding C=O stretching bands are reported in Figure 5. An upshift of the frequency of the peak position can be observed with the increase in stacking fraction. This observation suggests that the upshift observed in the experimental spectrum when moving from GMP to dG10 (see Figure 3A) is induced by the stacking interactions among the oligomer bases. This result is also in line with our previous observation in a single-stranded oligomer made of thymine bases, for which we observe that the relative intensity of the two C=O bands is modulated by stacking interactions.[21]

After computing the C=O vibrational modes of the selected DNA model systems made up of homogeneous single base components (MX, XMP and dX10, X=G or C), we move to a more complete and realistic structure. We model a seven-bases ss-DNA oligomer (7merDNA, 5'-AGGTTCA-3') featuring all the nucleobases and whose experimental spectrum is available.[24] The MD simulation was performed for 200 ns along which the unrealistic coil-like initial structure progressively evolves towards a more globular and collapsed assembly with increased pairing and stacking interactions, as shown in the time evolution of the root-mean-square deviation, the radius of gyration, the number of hydrogen bonds among the bases, and the stacking fraction in Figure S7 in the ESI. To exclude the initial portion of the MD trajectory where the structure is most likely to sample non-equilibrated conformers, we use the trajectory portion between 50 and 200 ns for the calculation of the 7merDNA IR spectrum, which is shown in Figure 6.

Since adenine does not include any C=O functional group and thymine has two C=O bands (C_2=O and C_4=O[21]) the spectrum of 7merDNA in the C=O stretching region includes seven bands: one band for each G and C base and two bands for each T base, see Figure 6A. In the experimental work,[24] the main signal that dominates the FTIR spectrum appears at 1668 cm^{-1} and is assigned to an overlapping contribution of C_4=O of free (i.e., non-base-paired) thymines, C=O of free guanines, and C=O of free cytosines. The high frequency minor shoulder at 1696 cm^{-1} is assigned to the C_2=O stretching of non-base-paired thymines. It can be observed in Figure 6A that the positions of the calculated spectral bands match the experimental assignments very well. However, the bands reported in panel A do not embed intra- and inter-modes excitonic couplings. In Figure 6B the total spectrum,

Physical Chemistry Chemical Physics Accepted Manuscript

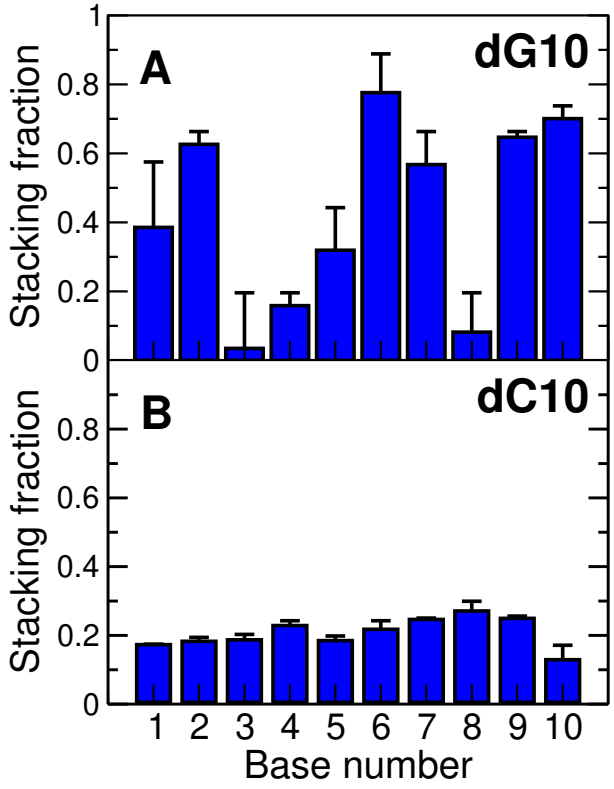


Fig. 4 Average stacking fraction for each base along the MD trajectories of dG10 (A) and dC10 (B).

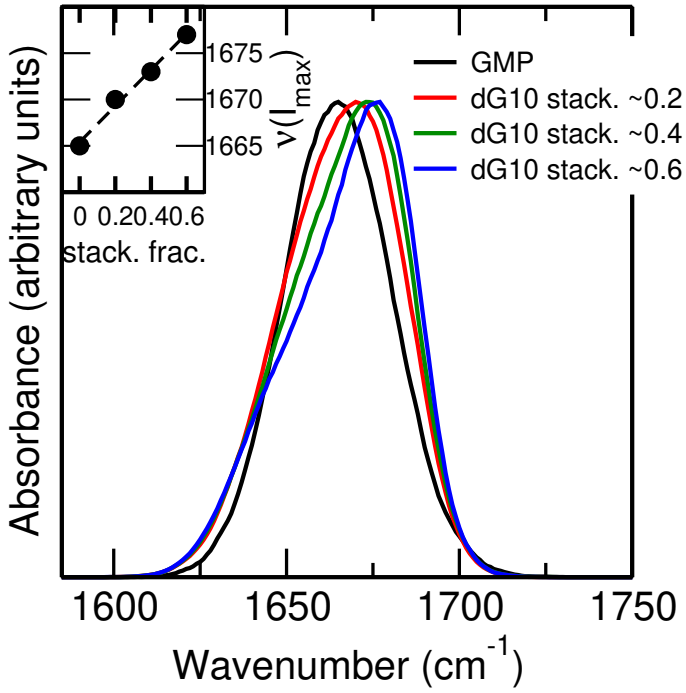


Fig. 5 Calculated spectrum of dG10 in D_2O at different stacking fraction conditions: red, ≈0.2 stacking fraction; green ≈0.4 stacking fraction; blue, ≈0.6 stacking fraction. In black the calculated spectrum of GMP is reported as null stacking fraction reference. In the inset, the peak frequency peak is reported for the four stacking fractions.

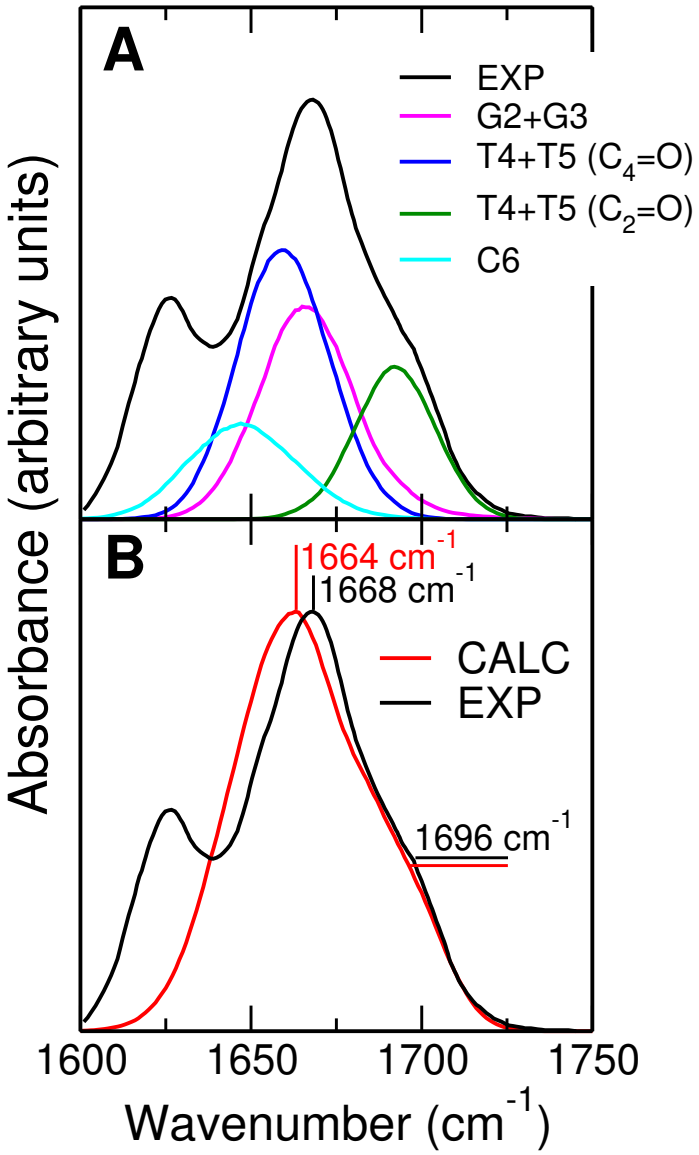


Fig. 6 A: Calculated signals arising from individual bases of 7merDNA oligomer. The legend reads as follows : magenta, G bases (G2+G3) ; blue, T bases (C_4=O modes, T4+T5); green, T bases (C_2=O modes, T4+T5); cyan, C base (C6); black, experimental spectrum. B: Experimental[24] (black) and calculated (red) spectra of 7merDNA with inclusion of inter-and intra-modes excitonic coupling. Note that the lower frequency band/shoulder presents in the experimental spectrum arises from the C=C or C=N stretching modes not accounted for in our calculations. The IR signals in B are normalized at the intensities of the highest peak.

with the inclusion of intra- and inter-mode excitonic couplings in the calculation, is compared to the experimental spectrum. It can be observed that all the features of the experimental IR band, including the spectral lineshape and peak frequencies, are very well reproduced in the calculated spectra. Specifically, the calculated frequency of the main peak (1664 cm^{-1}) is only slightly underestimated ($\Delta\nu = 4$ cm^{-1}), while that of the secondary shoulder (1696 cm^{-1}) perfectly matches the corresponding experimental value. The relative intensity of the main peak and secondary shoulder is also in excellent agreement with the experiment. As shown in Figure S8 in ESI, the inclusion of excitonic coupling ap-

Physical Chemistry Chemical Physics Accepted Manuscript

pears to be crucial for the proper reproduction of the experimental lineshape. In fact, without the inclusion of excitonic coupling, the main peak frequency is farther away from the experimental one ($\Delta\nu = 8$ cm^{-1}), and the secondary shoulder is not present.

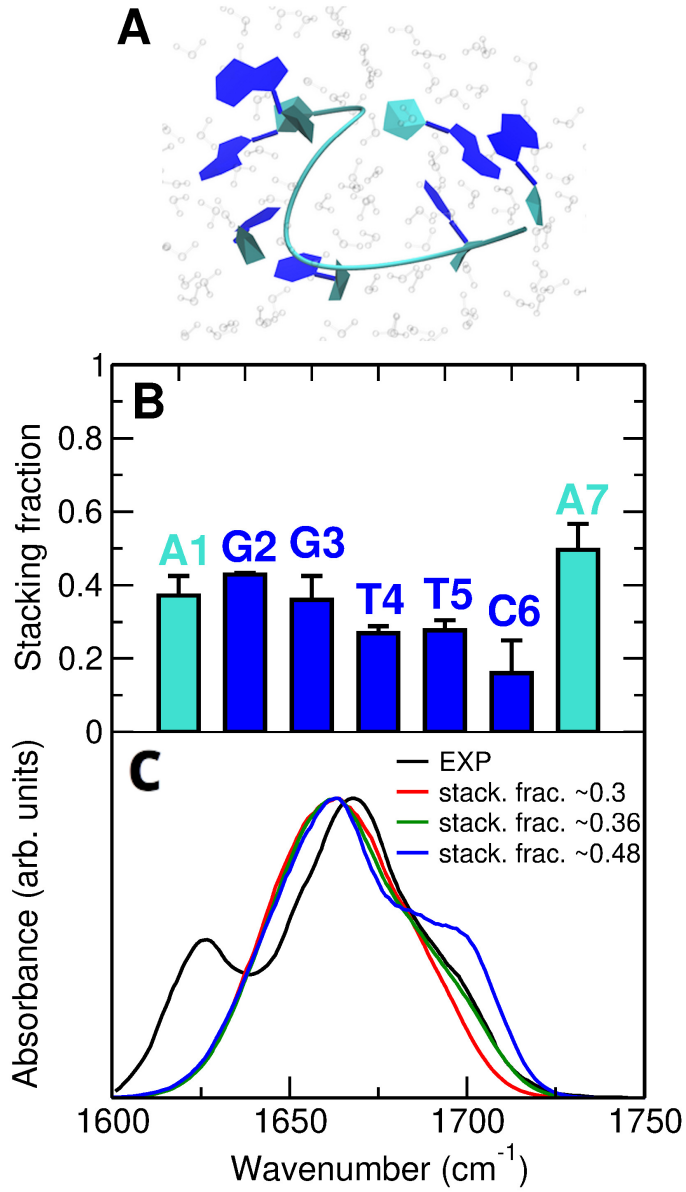


Fig. 7 A : Representative snapshots of the tmerDNA oligomer. Stacked bases are shown in blue and the first solvation shell is depicted by the grayscale water molecules. B: Stacking fraction of each base along 7merDNA MD trajectory. Adenine bases do not include a carbonyl C=O group but contribute to the stacking interaction among the bases with the corresponding contribution highlighted in cyan. C: Calculated spectrum of 7merDNA oligomer at different stacking fraction conditions: red, ≈0.3 stacking fraction; green ≈0.36 stacking fraction; ; blue ≈0.48 stacking fraction. The sequence of 7merDNA is reported on the top of panel A.

Our previous analyses for guanine and thymine[21] showed that the stacking fraction significantly affects the shape of the spectral lines. Therefore, we also analyze for the 7merDNA oligomer the sensitivity of the spectrum to stacking interactions. The average stacking fraction per-base in the 7merDNA oligomer is shown in Figure 7B. The corresponding time-related plot is reported in Figure S7C in the ESI. In agreement with our previous observations and with what is expected from pyrimidine and purine bases, a higher stacking fraction is observed in guanine and adenine (≈ 0.42) relative to cytosine and thymine (≈ 0.23). To further correlate the stacking fraction with the spectral features of 7merDNA, we divide its conformational ensemble into different subspaces using the stacking fraction as a metric. We then compute the spectrum of the 7merDNA oligomer using these subconformational spaces featuring different stacking fractions (Figure 7C). It turns out that at low stacking fraction (red line), the secondary shoulder almost disappears, while at high stacking fraction (blue line) the secondary shoulder is more pronounced, even cleaving to a separate peak at stacking fraction greater than or equal to 0.6 (data not shown). It should be noted that, in the final part of the trajectory (150-200 ns), the system adopts a relatively collapsed structure with significant stacking interactions (> 0.4), which appears as a stable conformation (see Figure S7 in the ESI). However, the spectrum corresponding to this sampling window (blue line in Figure 7C) does not match the experimental one. This result strongly suggests that, in the MD simulation, the coupling among the bases may be overestimated, likely pertaining to an over-representation of stacked configurations. This observation aligns with the interpretation provided in the experimental work suggesting a large predominance of free bases in 7merDNA.[24] The difference between the experimental and simulated conformational ensembles may partly arise from different ionic atmospheres, as only counterions required for electroneutrality were included in the simulations, without added salt to reproduce a specific ionic concentration. Furthermore, since classical force fields are mostly optimized to stabilize the structure of double-stranded DNA, an overestimation of stacking and/or pairing interactions in classical MDs of single-stranded DNA appears likely. An improved description of the conformational ensemble could be achieved with polarizable force fields, that have recently emerged as promising tools for nucleic acids modeling providing a balanced description of the stability of different forms.[61,62] The inclusion of polarizable force field sampling with the MD-PMM is technically possible within the current framework and could be pursued in future works.

To further assess the sensitivity of the spectral lineshape to the conformational ensemble of single-stranded DNA, we compute two additional geometrical descriptors, namely slide and twist, along the MD trajectories of dG10, dC10, and 7merDNA. Slide and twist are translational and rotational parameters, respectively, used to describe the relative arrangement of adjacent bases (details on their calculation are provided in section **??** of the ESI). As shown in Figure S9A in the ESI, the twist parameter is essentially homogeneous, featuring an approximately Gaussian distribution in all three cases, whereas the slide parameter displays greater variability, with its distribution suggesting the presence of distinct slide configurations along the MD trajectories (Figure S9B in the ESI). Calculation of the spectra for the different slide configurations (Figure S10 in the ESI) shows that,

Physical Chemistry Chemical Physics Accepted Manuscript

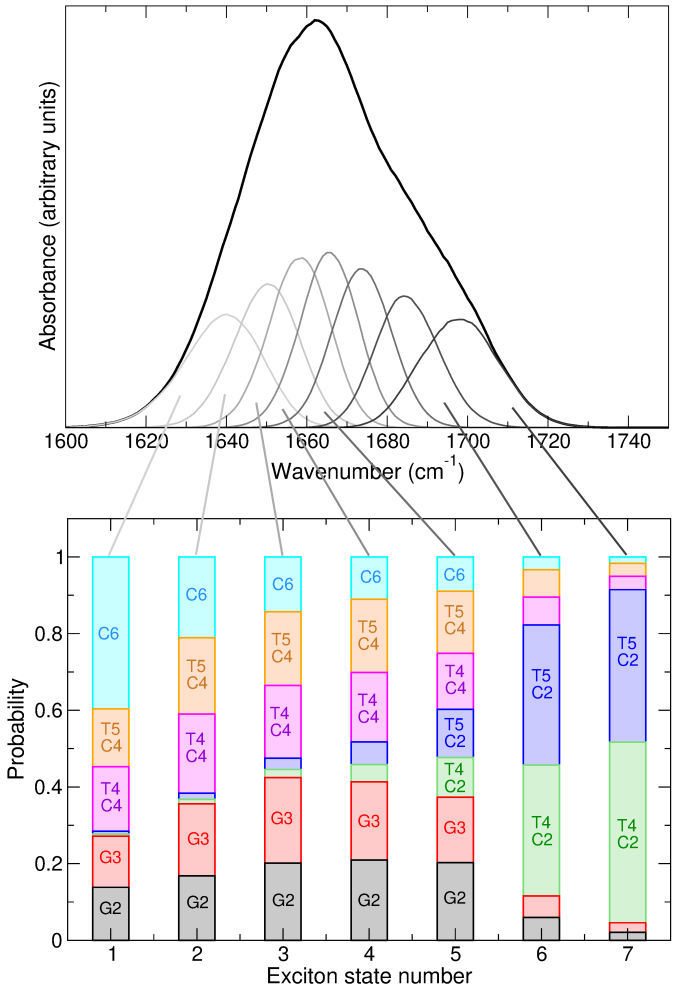


Fig. 8 Top: Calculated spectrum of 7merDNA in D_2O embedding excitonic modes coupling (black line). The individual spectra corresponding to the seven excitonic states 1-7, arising from the intra- and inter-modes of the 2 C=O of guanine, 2 C_2=O and C_4=O of thymine and 1 C=O of cytosine modes, are also reported. Bottom: The contributions of each mode to the excitonic states analyzed here are reported: guanine residues G2 and G3 in black and red, respectively; C_4=O stretching modes of thymine residues T4 and T5 in magenta and orange, respectively; C_2=O stretching modes of thymine residues T4 and T5 in green and blue, respectively; cytosine C6 in cyan.

for all three systems, the slide parameter does not significantly affect the overall spectral lineshape. Nonetheless, the features observed for different stacking configurations are also observed, albeit less pronounced, for the different slide configurations: the larger slide values in our trajectories, which roughly correspond to more stacked configurations, are associated with spectra in which the peak frequency is slightly upshifted for dG10 and the high-frequency shoulder is more pronounced for 7merDNA, consistently with the trends observed in Figures 5 and 7, respectively.

We then move to investigating the coupling effects among the vibrational modes in 7merDNA, to assess how much the individual carbonyl C=O stretching modes are coupled. We therefore analyze the contribution of each vibrational mode to each of the seven excitonic state spectra that build up the final spectrum. This can be done by averaging the square modulus of each component of each excitonic state expressed in the basis of the seven vibrational modes as obtained by the diagonalization of the excitonic Hamiltonian matrix along the MD trajectory. As depicted in Figure 8, the spectra arising from excitonic states 1 to 5 originate from the vibrational modes of the two guanines G2 and G3, cytosine C6 and from the C_4=O modes of the two thymines (T4_C4 and T5_C4). A minor contribution from the C_2=O modes of the two thymines (T4_C2 and T5_C2) is also present. The spectra arising from excitonic states 6 and 7 mainly arise from the highest frequency C_2=O modes of the thymine bases (T4_C2 and T5_C2). These results show that, in agreement with the experimental assignments, the low-frequency main peak essentially derives from the overlap between the vibrational modes of guanine and cytosine and from the C_4=O band of thymine, while the secondary minor shoulder mainly ($\approx$ 75%) results from the C_2=O mode of thymine. In addition, they also show that there is a strong interplay among vibrational modes, as evidenced by the mixed character of the excitonic states, each involving contributions from multiple normal modes. We also compute the inverse participation ratio, $IPR = 1/\sum_i |c_i|^4$, considered as a measure of the effective delocalization of the excitonic wave function on the interacting vibrational modes. Taking into account all seven excitonic states in 7merDNA, we got an estimate $IPR = 1.41 \pm 0.09$ (mean and standard deviation over all the excitonic states). This mean value is consistent with a conformational ensemble with limited interactions among the bases.

4 Conclusions

We present the calculation of infrared spectra arising from the C=O stretching modes of guanine- and cytosine-derived systems using a tandem theoretical-computational approach based on the joint use of classical molecular dynamics simulation, quantum mechanical calculation on the isolated base, and the perturbed matrix method (PMM). In the first instance, we select as a benchmark the MX model systems (X=G or C) that embed 9-methylguanine (9MG) and 1-methylcytosine (1MC). Subsequently, we moved to more extended and complete structures that include guanidine and cytosine 5'-monophosphate XMP model systems, and the single-stranded deca-homonucleotide of the guanine and cytosine dX10 conformers. For completeness, we also model a seven-base single-stranded DNA oligomer embedding all nucleobases including thymine already accounted for.[21] Overall, the calculated spectra well reproduce their experimental counterparts. Specifically, the calculations reproduce the frequency upshift between the experimental spectra of GMP and dG10, tracing its origin to intra-mode couplings among the bases due to stacking interactions. In contrast, no frequency shift is found between the calculated spectra of CMP and dC10, in accordance with the experimental data. For the single-stranded 7merDNA oligomer, the computed spectrum properly aligns to the experimental one, well reproducing the frequency position and relative intensity of the main peak and high frequency shoulder. The intensity of the latter was found to be extremely sensitive to stacking interactions among the bases, which induce excitonic coupling effects. Moreover, our results suggest that the conformational landscapes of the single-stranded 7merDNA oligomer

Physical Chemistry Chemical Physics Accepted Manuscript

is characterized by moderate stacking interactions and a modest network of hydrogen bonding interactions among the bases. Taken together, the results show that the MD-PMM methodology is able to properly capture the main spectral features of C=O stretching modes in DNA bases and, most importantly, the spectral changes induced by environmental effects.

Author contributions

SC: conceptualization, writing-review & editing, supervision, funding acquisition, project administration. CJDF: investigation, software, formal analysis, methodology, visualization, writing - original draft, writing - review & editing. GP: methodology, software, visualization, writing - original draft, writing - review & editing. LZP: conceptualization, methodology, formal analysis, investigation, writing - original draft, writing - review & editing, visualization, supervision, funding acquisition, project administration.

Conflicts of interest

There are no conflicts to declare.

Data availability

The data supporting this article are included as part of the ESI. Supplementary information: additional figures; output files from quantum mechanical calculations performed on both 9MG (9MG-D) and 1MC (1MC-D); initial structures, input parameters, and topology files for all molecular dynamics simulations.

Acknowledgments

The authors acknowledge support from the European Innovation Council through grant n. 101046920 "iSenseDNA". We acknowledge Lorenzo Dei Rossi for computing the IR spectra of 9MG and dG10 using the TIP4P/2005 water model. We acknowledge the CINECA awards HP10C1I45I and HP10C4UIDG for the availability of high-performance computing resources and support under the ISCRA initiative. The authors also acknowledge the use of the C3P (Computational Chemistry Community in Padua) HPC facility of the Department of Chemical Sciences of the University of Padua and the support by the high-performance computing infrastructure developed under the project "CONVECS", funded by the PR Veneto FESR 2021-2027 program, Priority 1 – Specific Objective 1.1 – Action 1.1.2. The network support from the COST Action CA17139 is acknowledged.

Notes and references

1 S. Ham, S. Cha, J.-H. Choi and M. Cho, *The Journal of chemical physics*, 2003, **119**, 1451–1461.
2 S. Ham and M. Cho, *The Journal of chemical physics*, 2003, **118**, 6915–6922.
3 J. Jeon, S. Yang, J.-H. Choi and M. Cho, *Accounts of chemical research*, 2009, **42**, 1280–1289.
4 C. Lee, K.-H. Park and M. Cho, *The Journal of chemical physics*, 2006, **125**, 114508.
5 C. Lee and M. Cho, *The Journal of chemical physics*, 2006, **125**, 114509.
6 C. Lee, K.-H. Park, J.-A. Kim, S. Hahn and M. Cho, *The Journal of chemical physics*, 2006, **125**, 114510.
7 C. Lee and M. Cho, *The Journal of chemical physics*, 2007, **126**, 145102.
8 J. R. Schmidt, S. A. Corcelli and J. L. Skinner, *The Journal of Chemical Physics*, 2004, **121**, 8887–8896.
9 Y. Jiang and L. Wang, *The Journal of Physical Chemistry B*, 2019, **123**, 5791–5804.
10 Y. Jiang and L. Wang, *The Journal of Chemical Physics*, 2020, **152**, 084114.
11 W. Meng, H.-C. Peng, Y. Liu, A. Stelling and L. Wang, *The Journal of Physical Chemistry B*, 2023, **127**, 2351–2361.
12 C. Qian, Y. Liu, W. Meng, Y. Jiang, S. Wang and L. Wang, *Journal of Chemical Theory and Computation*, 2024, **20**, 10080–10094.
13 M. Scherlo, D. Phillips, R. Künne, E. Ippoliti, K. Gerwert, C. Kötting, P. Carloni, A. S. J. S. Mey and T. Rudack, *The Journal of Physical Chemistry B*, 2025, **129**, 11652–11665.
14 L. Zanetti-Polzi, S. Del Galdo, I. Daidone, M. D'Abramo, V. Barone, M. Aschi and A. Amadei, *Physical Chemistry Chemical Physics*, 2018, **20**, 24369–24378.
15 L. Zanetti Polzi, A. Amadei, M. Aschi and I. Daidone, *Journal of the American Chemical Society*, 2011, **133**, 11414–11417.
16 L. Zanetti-Polzi, M. Aschi, A. Amadei and I. Daidone, *The Journal of Physical Chemistry B*, 2013, **117**, 12383–12390.
17 L. Zanetti-Polzi, C. M. Davis, M. Gruebele, R. B. Dyer, A. Amadei and I. Daidone, *FEBS letters*, 2017, **591**, 3265–3275.
18 C. M. Davis, L. Zanetti-Polzi, M. Gruebele, A. Amadei, R. B. Dyer and I. Daidone, *Chemical Science*, 2018, **9**, 9002–9011.
19 S. M. V. Pinto, N. Tasinato, V. Barone, A. Amadei, L. Zanetti-Polzi and I. Daidone, *Physical Chemistry Chemical Physics*, 2020, **22**, 3008–3016.
20 S. Pinto, N. Tasinato, V. Barone, L. Zanetti-Polzi and I. Daidone, *The Journal of Chemical Physics*, 2021, **154**, 084105.
21 G. Parolin, C. J. Dongmo Foumthuim, S. Corni and L. Zanetti-Polzi, *Phys. Chem. Chem. Phys.*, 2025, –.
22 J. Liquier and E. Taillandier, *Infrared Spectroscopy of Biomolecules*, Wiley-Liss, New York, 1996.
23 M. Banyay, M. Sarkar and A. Gräslund, *Biophysical chemistry*, 2003, **104**, 477–488.
24 M. Lindqvist, M. Sarkar, A. Winqvist, E. Rozners, R. Strömberg and A. Gräslund, *Biochemistry*, 2000, **39**, 1693–1701.
25 I. Daidone, M. Aschi, L. Zanetti-Polzi, A. Di Nola and A. Amadei, *Chemical Physics Letters*, 2010, **488**, 213–218.
26 A. Amadei, I. Daidone, L. Zanetti-Polzi and M. Aschi, *Theoretical Chemistry Accounts*, 2011, **129**, 31–43.
27 *European Journal of Biochemistry*, 1983, **131**, 9–15.
28 P. Hohenberg and W. Kohn, *Physical review*, 1964, **136**, B864.
29 W. Kohn and L. J. Sham, *Physical review*, 1965, **140**, A1133.
30 a. Petersson, A. Bennett, T. G. Tensfeldt, M. A. Al-Laham, W. A. Shirley and J. Mantzaris, *The Journal of chemical physics*, 1988, **89**, 2193–2218.

Physical Chemistry Chemical Physics Accepted Manuscript

31 G. Petersson and M. A. Al-Laham, *The Journal of chemical physics*, 1991, **94**, 6081–6090.

32 C. Lee, W. Yang and R. G. Parr, *Physical review B*, 1988, **37**, 785.

33 S. Grimme, S. Ehrlich and L. Goerigk, *Journal of Computational Chemistry*, 2011, **32**, 1456–1465.

34 M. E. Casida and M. Huix-Rotllant, *Annual review of physical chemistry*, 2012, **63**, 287–323.

35 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, *Gaussian~16 Revision C.01*, 2016, Gaussian Inc. Wallingford CT.

36 G. Sheina, S. Stepanian, E. Radchenko and Y. Blagoi, *Journal of Molecular Structure*, 1987, **158**, 275–292.

37 M. Szczesniak, M. Nowak and K. Szczepaniak, *Journal of Molecular Structure*, 1984, **115**, 221–224.

38 M. Szczesniak, J. Leszczynski and W. B. Person, *Journal of the American Chemical Society*, 1992, **114**, 2731–2733.

39 C. S. Peng, K. C. Jones and A. Tokmakoff, *Journal of the American Chemical Society*, 2011, **133**, 15650–15660.

40 B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, *Journal of Computational Chemistry*, 1983, **4**, 187–217.

41 A. D. MacKerell Jr., N. Banavali and N. Foloppe, *Biopolymers*, 2000, **56**, 257–265.

42 K. Hart, N. Foloppe, C. M. Baker, E. J. Denning, L. Nilsson and A. D. J. MacKerell, *Journal of Chemical Theory and Computation*, 2012, **8**, 348–362.

43 W. E. Reiher III, *PhD thesis*, Department of Chemistry, Harvard University, 1985.

44 E. Neria, S. Fischer and M. Karplus, *The Journal of Chemical Physics*, 1996, **105**, 1902–1921.

45 K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov *et al.*, *Journal of computational chemistry*, 2010, **31**, 671–690.

46 M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek and G. R. Hutchison, *Journal of Cheminformatics*, 2012, **4**, 1–17.

47 G. Bussi, D. Donadio and M. Parrinello, *The Journal of Chemical Physics*, 2007, **126**, 014101.

48 M. Parrinello and A. Rahman, *Phys. Rev. Lett.*, 1980, **45**, 1196–1199.

49 B. Hess, H. Bekker, H. J. C. Berendsen and J. G. E. M. Fraaije, *Journal of Computational Chemistry*, 1997, **18**, 1463–1472.

50 T. Darden, D. York and L. Pedersen, *The Journal of Chemical Physics*, 1993, **98**, 10089–10092.

51 U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *The Journal of Chemical Physics*, 1995, **103**, 8577–8593.

52 D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. C. Berendsen, *Journal of Computational Chemistry*, 2005, **26**, 1701–1718.

53 N. Michaud-Agrawal, E. J. Denning, T. B. Woolf and O. Beckstein, *Journal of Computational Chemistry*, 2011, **32**, 2319–2327.

54 K. B. Beć, J. Grabska and C. W. Huck, *Analytica chimica acta*, 2020, **1133**, 150–177.

55 V. C. Chamorro, C. Tempra and P. Jungwirth, *The Journal of Physical Chemistry B*, 2021, **125**, 4514–4519.

56 C. Tempra, V. C. Chamorro and P. Jungwirth, *The Journal of Physical Chemistry B*, 2023, **127**, 1138–1143.

57 G. Giubertoni, M. Bonn and S. Woutersen, *The Journal of Physical Chemistry B*, 2023, **127**, 8086–8094.

58 S. Jafilan, L. Klein, C. Hyun and J. Florián, *The Journal of Physical Chemistry B*, 2012, **116**, 3613–3618.

59 A. Taghavi, I. Riveros, D. J. Wales and I. Yildirim, *Journal of Chemical Theory and Computation*, 2022, **18**, 3637–3653.

60 A. Plumridge, S. P. Meisburger, K. Andresen and L. Pollack, *Nucleic Acids Research*, 2017, **45**, 3932–3943.

61 Z. Jing, C. Liu, S. Y. Cheng, R. Qi, B. D. Walker, J.-P. Piquemal and P. Ren, *Annual Review of biophysics*, 2019, **48**, 371–394.

62 A. M. Salsbury and J. A. Lemkul, *Current opinion in structural biology*, 2021, **67**, 9–17.

Physical Chemistry Chemical Physics Accepted Manuscript

Data availability statement

The data supporting this article are included as part of the Electronic Supplementary information: additional figures; output files from quantum mechanical calculations performed on both 9MG (9MGD) and 1MC (1MC-D); initial structures, input parameters, and topology files for all molecular dynamics simulations.

Physical Chemistry Chemical Physics Accepted Manuscript