Vibrational spectroscopy simulation of solvation effects on a G-quadruplex

Davide Moscato, Fabio Gabas, Riccardo Conte,* and Michele Ceotto*

Dipartimento di Chimica, Università degli Studi di Milano, via Golgi 19, 20133 Milano (Italy)

E-mail: riccardo.conte1@unimi.it; michele.ceotto@unimi.it

Abstract

It is commonly believed that solvation effects on the vibrational properties of a solute are easily accounted for by simple rules of thumbs, i.e. solvating a polar molecule in a polar medium has the only effect of red shifting all its spectroscopical features and, similarly, solvating a polar molecule in a non polar medium has the opposite effect. In this work we use theoretical vibrational spectroscopy at quasi-classical and quantum approximate semiclassical level to gain atomistic insights about solvent-solute interactions for 2'-deoxyguanosine and the G-quadruplex. We employ the quasi-classical trajectory method to include full anharmonicity into our calculated spectra, and then introduce quantum nuclear effects by means of divide-and-conquer semiclassical spectroscopy calculations. Solvation is treated explicitly leading to a good reproducibility of the available experimental data and reliable predictions when an experimental reference is missing.

Introduction

The nowadays well known double helix structure of DNA is widely regarded as one of the most important discoveries of the last century. This kind of structure is characterized by a well defined pattern of

interactions between pairs of nucleobases and it is able to preserve the entire genetic information of living creatures. It is also now known that some nucleobases are able to interact with each other in a different way than the one described by Watson and Crick (WC).[1] To emphasize this concept, some computational studies demonstrated that guanine seems particularly prone to interact with other guanine molecules.[2–4] In many living organisms we are indeed able to observe this behavior, for instance in the guanine rich sequences found in the telomeric and promoter regions of DNA.[5] It is likely that these sequences fold into a non canonical double-helix WC structure, where four guanines interacting through Hoogsteen-type bonding are arranged in a square planar geometry. This type of geometries that are formed by one, two or four DNA strands, are also known as G-quadruplexes. Factors like the sequence of nucleobases, presence of certain ligands and other environmental effects determine the way the quadruplex will fold. For this reason many shapes of this kind of molecule can be found. Recently a lot of attention has been payed by the scientific community to these kinds of systems due to their potential functional role in gene regulation and DNA repairing mechanisms from lesions.[6,7]

The interest in G-quadruplexes comes mainly from their potential applicability as targets of cancer therapeutics,[8–12] but also for their possible role in virus replication activity for viruses like HIV and SARS-CoV-2.[13–16] This is due to the fact that G-quadruplexes serve as kind of "knot" in the DNA replication process signaling the stop of transcription. This means that the development of ligands stabilizing the quadruplexes could be used to induce the apoptosis of target cells.[17] Many efforts have been put also in the *in vitro* detection of specific G-quadruplexes by means of UV spectroscopy and Circular Dichroism (CD).[18,19] Other techniques used for the analysis of the quadruplexes in water solutions, which are also good candidates for the *in vivo* detection, are Raman spectroscopy (RS) methods such as surface enhanced Raman spectroscopy (SERS), polarized Raman scattering, and UV resonant Raman (UVRR).[20–22] While all of these approaches focus mainly on the effects of topology change, change of the templating cation and signal identification, none of them seem to point towards solvent effects, which are fundamental to be taken into consideration before moving to the in vivo analysis.

In this work our main goal is to understand the effects of solvation on the vibrational features of

the G-quadruplex. In particular, our work focuses on the differences that occur between gas phase spectroscopic simulations and simulations carried out in a box of solvent. The latter should also allow us to demonstrate that use of an appropriate explicit solvent is fundamental to catch the correct shifts caused by system-solvent interactions.

Vibrational spectroscopy is a useful tool often employed by experimental chemists and physicists to understand the structure and dynamics of molecules. Techniques such as the Fourier transform IR (FTIR) or more recently the IR multi photon dissociation (IRMPD) are able to give many important insights about molecules in gas phase, solid state and, in some cases, even liquid state.[23–33] However, their applicability to biological systems still remains problematic. This is due to the fact that the absorption of liquid water in the IR is very intense, causing the spectrum to be entirely dominated by the solvent features instead of those of the system. RS is ideal for this kind of applications, since liquid water has a low Raman activity. For these reasons RS has found large application to both chemistry, biochemistry and also the biotechnological field.[21,34] However, while all these techniques are able to give useful information about structure, dynamics and presence of a given molecular species, they are not able to catch the atomic details for any of these.

In the last years computer simulations have served as a tool to interpret and predict experiments. The work in our group has been mostly focused on the development and application of path-integral-derived semiclassical techniques for the simulation of vibrational spectroscopy[35] at the (approximate) quantum mechanical level. In particular, very recently, we have also shown that application of our methods to large systems is possible even adopting empirical force fields.[36,37]

In this paper we focus on biological systems. We start with a simple system like the 2'-deoxyguanosine, which we study both in gas and solvated phase. We adopt the information gained from this first set of simulations to show how solvation acts differently when the same molecule is found in a quadruplex arrangement, and compare our results with experimental data.

Theoretical and computational details

The reported simulations are characterized by more than 20000 degrees of freedom. Therefore, the only viable way to describe the potential energy surface (PES) of the system is the adoption of classical force fields. (FF) In a previous work presented by three of us,[36] we benchmarked several FF against ab initio data, and the polarizable FF AMOEBABIO18 included in the Tinker 8.6.1 software suite emerged as the most effective one.[38,39] An important point to make while using FF applied to spectroscopic simulations is that as the quality of the PES description is limited, so it is expected to be the accuracy with respect to experimental data. However, a qualitative and insightful description of the spectroscopic features is still achievable. For every studied system, a preliminary geometry optimization is carried out in gas-phase conditions, with a quasi-Newton algorithm. Then, the system is inserted in a solvent box with periodic boundary conditions (PBC), and particle mesh Ewald (PME) scheme for the long range interactions is adopted. Choice of the dimension and shape of the box is specific for the system and will be specified later in this section. Once the molecule is inserted in the PBC box, a second optimization with the steepest descent method is carried out. The molecular dynamics (MD) simulations are propagated for 3000 steps with a time step of 0.2 fs by using the Beeman integrator.[40] A particular set of initial mass-scaled conditions is used for the NVE trajectories

$$
\begin{cases}
p_j(0) = \sqrt{\hbar \omega_j (2n_j + 1)} \\
q_j(0) = q_{j,eq}
\end{cases}
,
\tag{1}
$$

where the initial position for the $j-$th normal mode is the one found for the equilibrium configuration, the initial linear momentum is obtained from the harmonic oscillator (HO) approximation, ω_j is the harmonic frequency for the $j-$th mode and n_j is a positive integer number which represents the quantum of excitation for the $j-$th mode. With the initial conditions shown in Eq. 1, it is possible to run trajectories at an energy close to the harmonic zero point energy (ZPE). By adding a quantum of excitation on a given mode, the energy of the trajectory will be closer to the energy of the quantum state corresponding to the excitation in that mode.

Quasiclassical trajectory

Most of the spectroscopic simulations we present are made at the quasi classical trajectory (QCT) level, where the initial conditions of the trajectory are quantized (equation 1) and the evolution of the system is purely classical. In this framework the molecular power spectrum is calculated for each vibrational mode j as the Fourier transform (FT) of the classical momentum (or velocity) autocorrelation function.

$$I_j(\omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \langle p_j(0)p_j(t) \rangle \ e^{i\omega t} \ dt \tag{2}$$

In QCT spectra, nuclear quantum effects are not accounted for, meaning that the spectral peaks correspond only to classical vibrational frequencies. Our approach with this method is to run a single trajectory with a quantum of excitation in the mode of interest. Usually a trajectory per mode of interest is sufficient to obtain a clean and accurate enough spectrum, but as it will be shown in the results section, for some modes a wider area of the phase space needs to be sampled to achieve a better qualitative description of the spectroscopic feature. In the latter instance, we use a set of trajectories with randomized initial mass-scaled conditions starting from those in Eq. 1.

$$\begin{cases} p_j(0) = -\sqrt{\hbar\omega_j(2n_j + 1)} \ \sin(\theta_j) \\ q_j(0) = \sqrt{\frac{\hbar}{\omega_j}(2n_j + 1)} \ \cos(\theta_j) \end{cases}, \tag{3}$$

where θ_j is a random uniformly distributed phase between 0 and 2π, n_j is set equal to 1 for the modes of interest and equal to 0 for all other modes. The resulting spectrum when several MD runs are performed is simply the average of the single spectra, i.e.

$$\langle I(\omega) \rangle_j = \frac{1}{N_{traj}} \sum_{i=1}^{N_{traj}} I_j^{(i)}(\omega) \tag{4}$$

where $I_j^{(i)}(\omega)$ is the intensity at frequency ω for the j-th normal mode given by the i-esime trajectory.

Semiclassical method

Semiclassical Initial Value Representation (SCIVR) has proved to be a practical and accurate tool for adding quantum effects to molecular dynamics.[41–58] Several semiclassical simulations are carried out also by means of the multiple coherent divide-and-conquer semiclassical initial value representation method (MC-DC SCIVR).[59–66] With this method it is possible to compute quantum mechanical vibrational power spectra of high dimensional systems upon projection of the dynamical quantities onto lower dimensional vibrational subspaces. In this case the working formula is

$$\tilde{I}(\tilde{E}) = \frac{1}{2\pi\hbar T}\frac{1}{N_{traj}}\sum_{i=1}^{N_{traj}}\left|\int_0^T e^{\frac{i}{\hbar}\left[\tilde{S}_t\left(\tilde{\mathbf{p}}_i(0),\tilde{\mathbf{q}}_i(0)\right)+\tilde{E}t+\tilde{\phi}_t\right]}\left\langle\tilde{\Psi}\Big|\tilde{\mathbf{p}}_i(t),\tilde{\mathbf{q}}_i(t)\right\rangle dt\right|^2, \tag{5}$$

where $\left(\tilde{\mathbf{p}}_i(0),\tilde{\mathbf{q}}_i(0)\right)$ is the initial phase space configuration projected on the subspace for the $i-$th trajectory, $\tilde{S}_t$ is the projected classical action evaluated along the trajectory, $\tilde{E}$ is a parameter and $\tilde{\phi}_t$ is the phase of the projected prefactor, which is defined as

$$\tilde{\phi}_t = \text{phase}\left[\sqrt{\left|\frac{1}{2}\left(\frac{\partial\tilde{\mathbf{q}}(t)}{\partial\tilde{\mathbf{q}}(0)} + \Gamma^{-1}\frac{\partial\tilde{\mathbf{p}}(t)}{\partial\tilde{\mathbf{p}}(0)}\Gamma + i\hbar\Gamma\frac{\partial\tilde{\mathbf{q}}(t)}{\partial\tilde{\mathbf{p}}(0)} + \frac{i\Gamma^{-1}}{\hbar}\frac{\partial\tilde{\mathbf{p}}(t)}{\partial\tilde{\mathbf{q}}(0)}\right)\right|}\right] \tag{6}$$

To evaluate the sub blocks of the monodromy matrix which appear in ϕ_t it is necessary to calculate the Hessian matrix at every time step of the MD simulation. This represents a huge bottleneck for the whole method. The problem of Hessian calculations has been tackled with different methods in our group.[67–69] Ab initio methods applied to molecules with more than 50 atoms can be seriously problematic under this aspect, while when using an FF the Hessian matrix can be computed at every step without approximations even for molecules with more than 100 atoms. For systems bigger than 1000 atoms, we opt to propagate along the trajectory only the elements of the Hessian related to the normal modes belonging to the subspace of interest. These elements of the Hessian matrix are approximated by differentiation of the forces calculated along the dynamics.[70,71] With this method we are able to focus only on the normal modes centered on the molecule and eliminate from the spectrum those coming from the solvent molecules. For SC calculations, a single trajectory has been used upon setting to zero

the initial momentum of all the normal modes with harmonic frequencies lower than 300 cm^{-1}, and providing all other normal modes with the basic zero-point energy contribution (i.e. n=0 in Eq.1).

Computational details

As anticipated in the introduction, we focus our attention on the study of the 2'-deoxyguanosine, and a tetrad of 2'-deoxyguanosine which from now on will be simply referred to as the "quadruplex". Guess geometries for both systems are extrapolated by removing the necessary residues from a DFT optimized structure published in a paper by Sponer et. al.[3] Then, minimization is performed for a single deoxyguanosine inserted in a 30 Å-side cubic box filled with water (2702 atoms), and the quadruplex inserted in a rectangular box with a 50×50 Å^2 base, and 30 Å height. (7517 atoms) For both systems long range interactions are considered using PME.[72,73] As a comparison with experimental data, we considered the Raman spectrum in water of the thrombin binding aptamer (TBA) by ref.[20] In this case we made the assumption that the quadruplex considered in our simulations can be thought of as a minimal sub-unit of the full TBA. Other experiments with other types of G-quadruplexes exist as well, but we think that since TBA is one of the smallest systems that can be found in literature it could better resemble the behavior of the minimal unit that we are considering. We are not currently aware of how much the presence of a rigid backbone like the one present in the TBA can influence our discussion, but this factor will be taken into consideration in the result discussion part. The vibrational normal modes considered in this work, for both studied systems are the asymmetric and symmetric stretches of the $C2'H_2$ group of the sugar moiety, the $C6 = O$ stretch, and the 2'-deoxyguanosine (dG) ring deformation mode.
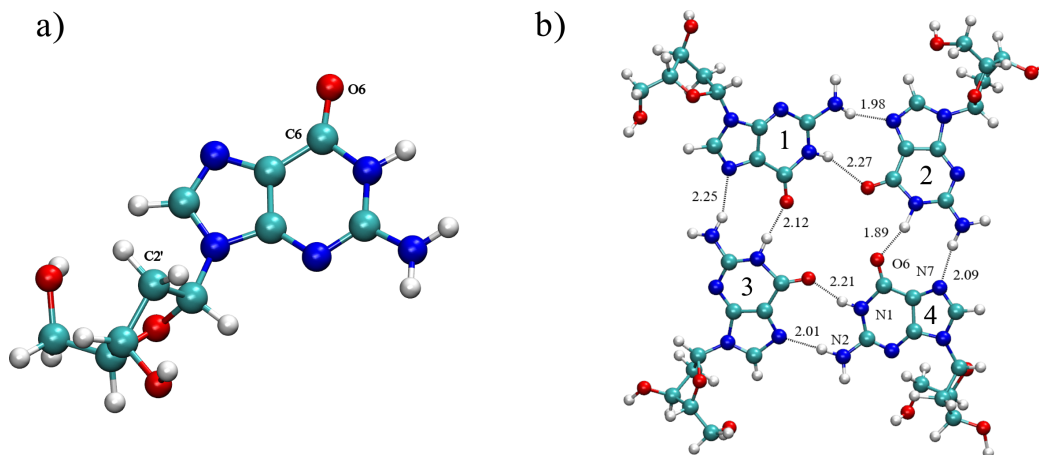
Figure 1: a) 2'-deoxyguanosine equilibrium geometry in the water box; b) quadruplex minimum geometry in the water box.

Results

2'-Deoxyguanosine

In the first set of QCT simulations performed on the 2'-deoxyguanosine molecule, we observe that solvation acts differently on different normal modes. Specifically, it is possible to observe that for the $C = O$ stretch, (figure 2, panel a) multiple vibrational features are present in both solvated and gas phase systems. In order to assign the $C = O$ stretch to the correct peak, a more detailed analysis has been conducted. For both systems two additional trajectories are run. In one all the initial momenta are set to the harmonic ZPE (ZPE trajectory). In the other one the initial momentum of the $C = O$ stretching mode is set to zero, while all the other modes have the same initial momenta as in the ZPE trajectory (Deactivated trajectory). By doing so, in the gas phase spectrum (figure s1 panel a in the supplementary information file) we observe for the deactivated trajectory a single peak much more intense than the one observed in the original trajectory, which we recall was characterized by a harmonic quantum of excitation in the stretch and the corresponding harmonic zero-point energy contribution in all other modes. The position of this peak is ≈ 30 cm^{-1} red-shifted with respect to the most intense peak in Figure 2 a. In the ZPE trajectory the same feature of the spectrum in Figure 2 a is observed, even if it is slightly blue-shifted and with a change in the intensity ratio between the most intense peak

and the secondary one. In particular it seems that the secondary one gains in intensity with these set of initial conditions, while the main one loses intensity. Based on these observations we can conclude that the most intense peak can be assigned to the C=O stretch because in the deactivated trajectory it provides an almost harmonic signal. Moving on to the analysis of the solvated system, (figure s1 panel b in the SI file) we observe that both the deactivated and ZPE spectra are much less intense than the one in Figure 2 a. In the spectrum generated by the deactivated trajectory, a single peak is again observed. It is positioned almost at the same frequency of the secondary peak on the right of the main one. A similar situation is observed with the ZPE trajectory, with the main difference that the observed feature is enlarged with a shoulder near the position of the most intense peak of the original spectrum. (figure 2 a) Again, these two observations led us to the conclusion that the $C = O$ stretch is the one positioned at 1553 cm^{-1}, which corresponds to the most intense one. With this we can conclude that a red shift is observed when solvent molecules are added, on the contrary of the harmonic estimates.

By a simple visual analysis of the MD simulation whereby the spectrum has been calculated, it is possible to observe that this part of the molecule is strongly interacting with the surrounding water molecules. We think that this interaction with water is what causes the observed red shift of the QCT signal with respect to the gas phase case. Another peculiarity that can be observed from the spectrum in figure 2, is that while the QCT signal is reported to be strongly harmonic in gas phase, in the water box simulation the harmonic estimate is 115 cm^{-1} far from the QCT signal. Probably this anharmonicity derives from the fact that in the equilibrium conformation, from which the harmonic estimate is made, the $C = O$ group sense some of the interactions with the solvent as repulsive. This could explain also why the harmonic estimates of the CO stretch in the water box system are blue shifted with respect to the harmonic estimate made in vacuo. This result demonstrates the importance of anharmonicity and dynamical effects when simulating biomolecules in solution.

The opposite effect was observed for the ring deformation mode (dG RDM). In this case solvation seems to blue shift both the QCT and harmonic estimates with respect to the signal observed in gas phase. (fig. 2 panel b)

DG RDM is a vibrational normal mode delocalized on all atoms of the guanine. For this reason a
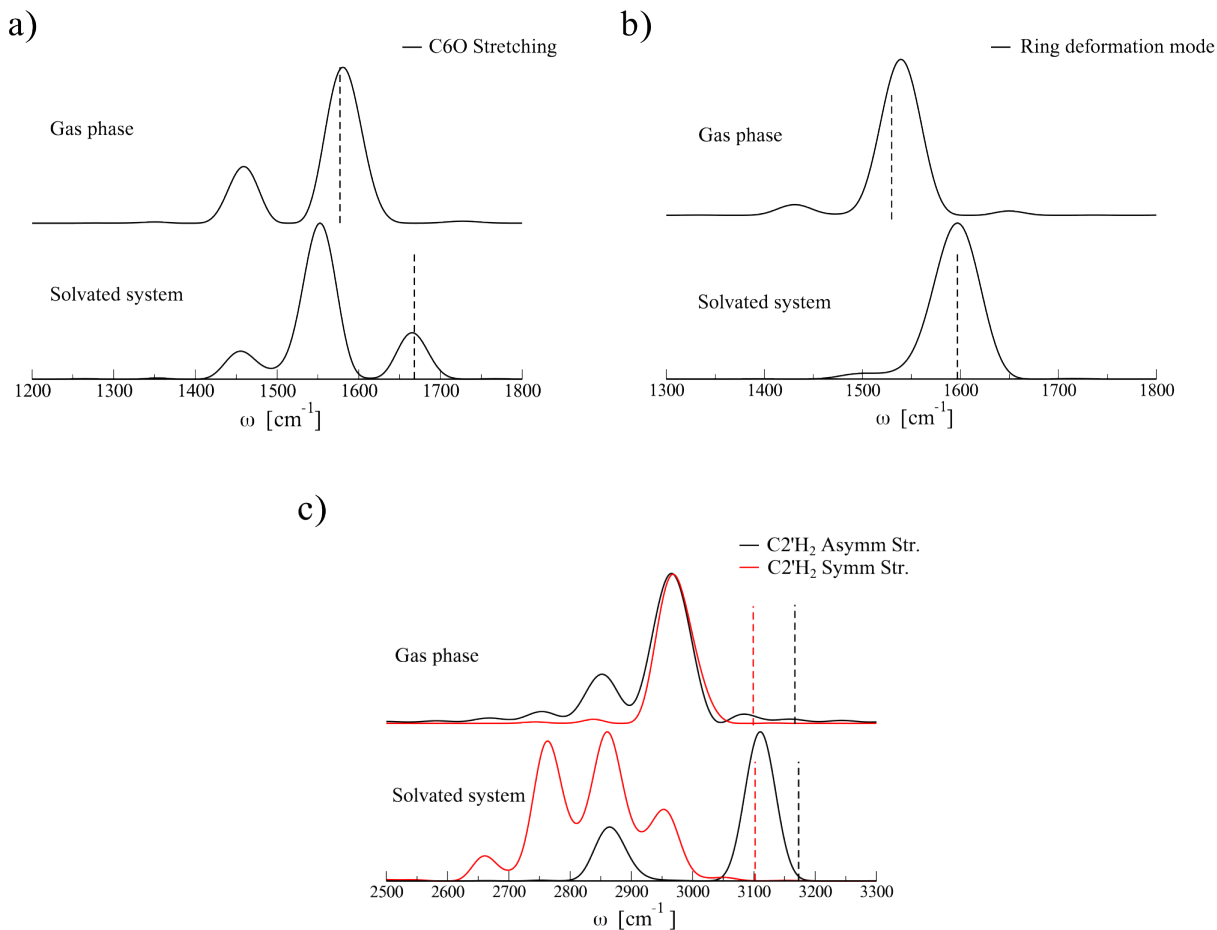
Figure 2: 2'-deoxyguanosine QCT spectra with a single trajectory per normal mode. Panel a) $C = O$ stretch, b) ring deformation and c) asymmetric and symmetric stretches of $C2'H_2$. Vertical dashed lines indicates the harmonic estimates.
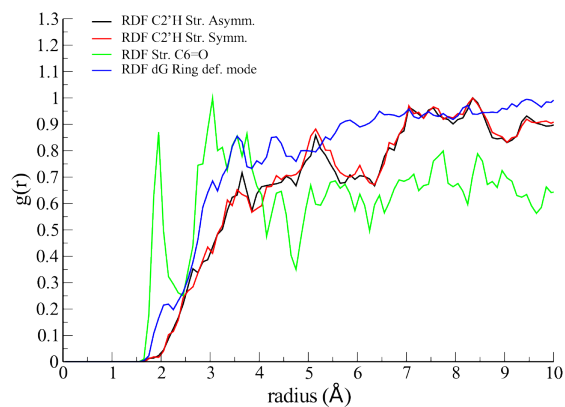

Figure 3: 2'-deoxyguanosine Radial distribution function (RDF) for the trajectory associated to the analyzed normal modes calculated between the group of atoms of interest and water.

simple visual analysis of the MD simulation may be not sufficient to understand why the spectrum in figure 2 (panel b) features blue shifted QCT and harmonic frequencies with respect to the corresponding ones in the gas phase simulation. By calculating an average[74] of the radial pair distribution function (RDF) between all the atoms of the guanine ring and all that ones of the water molecules (blue solid line in figure 3), it is possible to observe that water tends to occupy outer solvation shells, except for a little peak near 2 Å, which can be interpreted as the presence of labile short range interactions. This probably means that the interactions between water and the whole guanine ring tend to be unfavorable. By observing the RDF for $C6 = O$ (green solid line in figure 3) atoms, it can also be found the proof that during the MD simulation this part of the molecule is strongly interacting with water. In particular this can be deduced by the fact that a high peak is present at distances shorter than the first solvation shell.

The last set of normal modes that we analyze for the 2'-deoxyguanosine is localized on the $C2'H_2$ group on the sugar moiety of the molecule.

By observing the spectrum of the $C2'H_2$ modes, (figure 2 panel c) it is expected to be difficult to extract much information. This is due to the fact that in the QCT gas phase spectrum associated to the trajectory of the symmetric stretch there is also a signal related to the asymmetric stretch, while for a water solvated system the signals are split but do not provide a clear fingerprint. For the gas phase system, we relate the presence of several peaks to the fact that a single trajectory for these two modes is not sufficient to sample a wide enough phase-space volume around the minimum geometry.

For this reason we move to a multiple trajectory approach, with initial conditions as stated in equation 3. In particular, we try this method with, 10, 100 and 1000 trajectories for the gas phase system, for all the normal modes to be studied.

By observing the spectra in figure 4, it is possible to see that by increasing the number of trajectories from 10 to 1000 a better resolution of the separation between the QCT asymmetric and symmetric stretches for $C2'H$ is obtained. On the other hand, for the $C = O$ stretch and the ring deformation mode (figure s2 in the SI file) a reduced number of trajectories is already sufficient to provide a properly converged signal. This could be due to the fact that normal modes like the $C = O$ stretches and dG
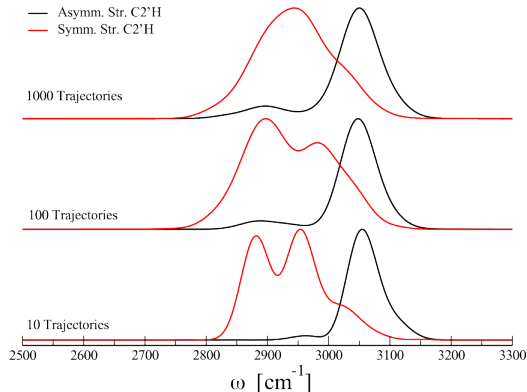
11

Figure 4: $C2'H_2$ stretches QCT spectra averaged on multiple trajectories for the gas phase simulations of 2'-deoxyguanosine.

RDM are usually rather harmonic, meaning that a more restricted volume of phase space can be sampled to recover the correct spetroscopic feature.
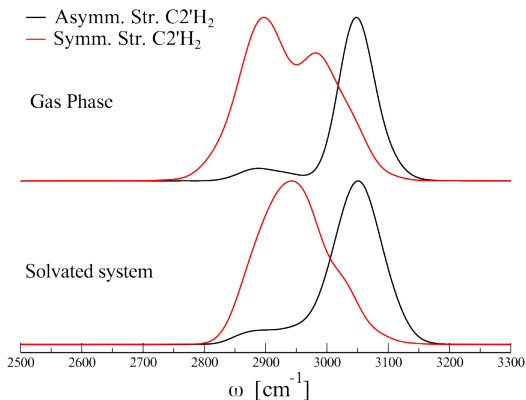

Figure 5: 2'-deoxyguanosine QCT spectra of the $C2'H_2$ modes averaged on 100 trajectories.

Given these results we are now ready to apply our multi-trajectory approach to water solvated systems. We do this by employing 100 trajectories given the unaffordable overhead of adopting 1000 trajectories for such systems. By comparing gas-phase and "in solution" spectra (figure 5) we are eventually able to assess the effects of solvation on the $C2'H_2$ normal modes. We find that results are very similar in the two different phases. This makes us believe that water molecules have no sizable interactions with the $C2'H_2$ group. This can be verified by looking at the RDF plots for the two trajectories in question (fig. 3 black and red lines). In these cases it can be observed that water tends to be organized in solvation

12

shells without any sign of short range interaction.

Quadruplex

In our vibrational analysis of the single 2'-deoxyguanosine molecule we observe that the effects of solvation are quite complex and not easily detectable by adopting simple rules of thumb. Moving on to a more complex system like the quadruplex (figure 1b) will further prove this concept true. In particular, in a system like the quadruplex it is possible to ascertain that a well-defined pattern of interactions between the four 2'-deoxyguanosine residues is present. We will confirm if the water molecules are able to interact mainly with the external part of the system, while the core of the system is partially hindered to interact with the solvent. More than that, in a structure like the one of the quadruplex, we expect that the vibrational frequencies of the modes under investigation stick to a symmetry related order. The whole supramolecolar system which composes the quadruplex (figure 1) should ideally belong to the D_{4h} symmetry group. Therefore, in an ideal scenario, the normal modes which we are considering should reveal a 4-fold degeneracy. In reality there are small deviations from this symmetric arrangement and so symmetry-related frequencies are expected to be close to each other but not exactly degenerate.

In the experimental spectrum of TBA,[20] the signals of the $C2'H_2$ stretches are assigned to a broad low intensity feature. The main peak at 2952 cm^{-1} is related to the asymmetric stretch, while the shoulder at 2890 cm^{-1} to the symmetric stretch. The $C = O$ stretches are experimentally assigned to the intense and broad peak centered at 1663 cm^{-1}, while the dG RDM mode is assigned to an intense narrow peak at 1582 cm^{-1}.

We start our vibrational analysis of the quadruplex by looking at the $C2'H_2$ normal modes. As it was observed on the previous case, a single trajectory per normal mode is not sufficient to extract fully reliable information. (figure s3 in the SI file)

For this reason we calculate the spectrum by running 100 trajectories for both the gas and solvated system. With this approach we are able to reproduce the experimental feature. On the other hand, our results do not have a strong quantitative agreement with respect to the experimental data, resulting in

a Mean Absolute Error (MAE) of 107 cm^{-1}. As anticipated in the Introduction, this is mainly due to the force field we need to employ to describe the PES of the system.
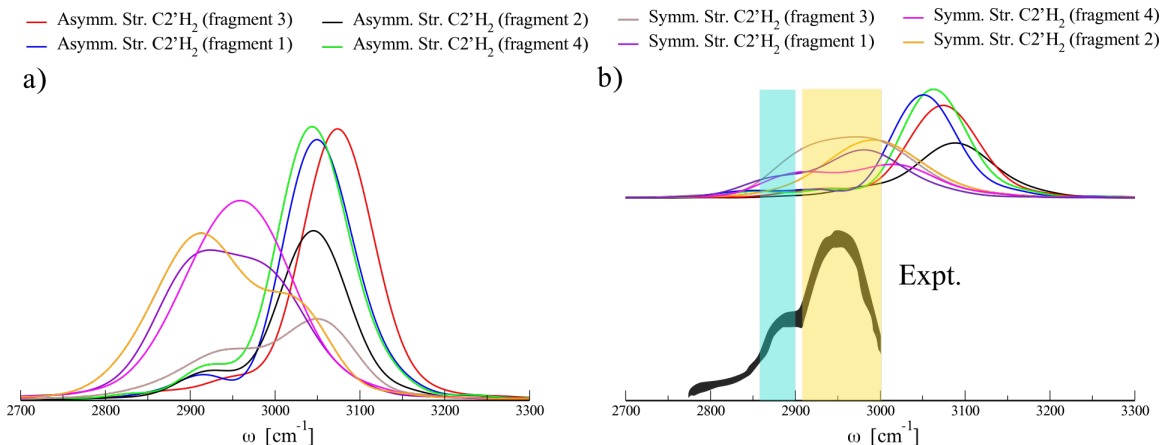

Figure 6: Gas phase a) and solvated b) $C2'H_2$ stretches QCT spectra calculated by means of 100 trajectories for the quadruplex. The profile of the experimental feature is presented as a broad black line, where the asymmetric streching is highlighted in orange and the asymmetric one in teal.

On the basis of the calculated spectra in figure 6, it is also easy to see that even in this case the solvation of the system has nearly no effect on the $C2'H_2$ signal positions. In the case of a single 2'-deoxyguanosine, this part of the molecule seems to have almost no interaction with the water molecules and for the quadruplex we think that this should not change much. In TBA, the tetrad of guanine is surrounded by a rigid backbone and the region near the $C2'$ is excluded from interacting with water as well, and, for this reason, we think that considering a model where this part of the molecule is weakly interacting with water should not be a rough approximation.

Moving on to lower frequencies, we analyze the $C = O$ stretches. In our optimized structure of the quadruplex, water molecules are able to interact with the $C = O$ group of the quadruplex only from above or below and this makes this part of the molecule less prone to solvent interaction.

By observing the spectrum in figure 7, it is clear that in this case the addition of solvent molecules to the simulation has the effect to blue shift the frequencies with respect to the gas-phase simulation. The opposite effect is observed for the single 2'-deoxyguanosine molecule, which means that solvation has a different effect on the vibrational features of the quadruplex. We use again the RDF as a tool to better understand our results. (figure s4 in the SI file) In this case we find out that the shape of the function
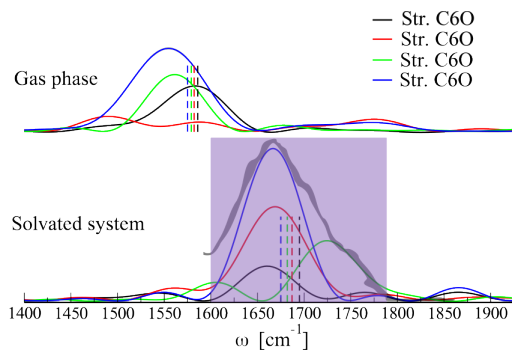
14

Figure 7: Quadruplex $C6O$ stretches QCT spectrum (solid colored lines) and the relative harmonic estimates (vertical dotted lines) for the gas-phase and solvated quadruplex.

is much similar to the one of the dG RDM mode reported in figure 3 for the single 2'-deoxyguanosine. This means that the interaction of the $C = O$ group with water is weak and unfavorable. The agreement with the experiment for the $C = O$ stretches is good, and the QCT signals for these modes computed for the quadruplex are also blue shifted with respect to the one computed for the single 2'-deoxyguanosine. This shift, as reported in the literature,[75] is a distinctive sign of the presence of the 2'-deoxyguanosine in a G-quadruplex arrangement instead of a single/double strand configuration. And can also be observed in our computed data (table 1)

The last set of normal modes that are left to be examined are the dG RDM modes. Our observations in this case are mostly aligned with what was observed with the single 2'-deoxyguanosine, i.e. this set of modes are strongly harmonic, and solvation makes both the QCT and harmonic estimates blue shift. The main difference lies in the reason of the blue shift, since the water molecules are not able to get as close to the guanosine rings as in 2'-deoxyguanosine.

Likewise the case of the $C = O$ stretches of the quadruplex, in the dG RDM spectrum (figure 8) we observe that both the harmonic and QCT signals are in good agreement with experimental data. The main difference between the harmonic estimates and the QCT frequencies is that the latter are nearly degenerate like the experimental ones. As anticipated, our simulation concerns only a minimal unit of the full TBA. This means that compensation of errors could be a factor in our last set of results, but it has to be noted that, like in the full structure, also the core part of the quadruplex has low capability of interaction with water molecules. Therefore it is possible to suggest that the effect of the presence

15

Table 1: QCT frequencies of the studied modes compared between the single 2'-deoxyguanosine and the quadruplex and to experimental data of TBA from ref. 20

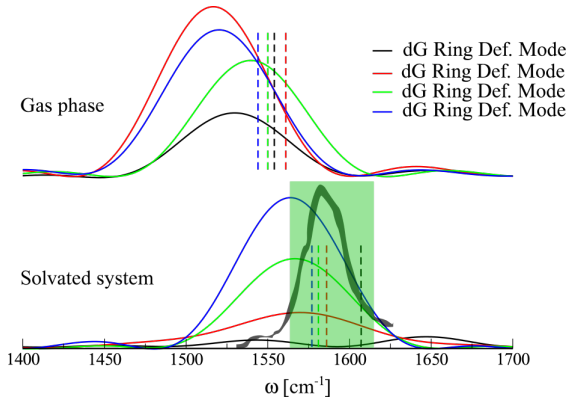| | Summary table of QCT frequencies (cm^{-1}) | | | | |
| | 2-deoxyguanosine | | quadruplex | | |
	gas phase	water box	gas phase	water box	exp. TBA
Asymm. Str. $C2'H_2$ (100 trajectories)	3048	3051	3074 3049 3045 3044	3088 3074 3051 3051	2952
Symm. Str. $C2'H_2$ (100 trajectories)	2897	2942	3049 2924 2959 2913	2992 2972 3014 2981	2890
Str. $C6 = O$ (1 trajectory per mode)	1581	1553	1583 1587 1561 1555	1660 1669 1724 1667	1663
dG RDM (1trajectory per mode)	1540	1568	1530 1517 1540 1520	1542 1570 1567 1564	1582
				MAE	63 cm^{-1}

Figure 8: Quadruplex dGRDM QCT spectrum (solid colored lines) and the relative harmonic estimates (vertical dotted lines) for the ring deformation modes. The green band identifies the experiment.

of a rigid backbone is to preserve the core part of the G-quadruplex, which means that a structure without it opens the possibility on longer time scales to see a dissociation of the supramolecular system. Following this argument we think that the role of the water molecules in the quadruplex simulations is to constrain the four fragments of 2'-deoxyguanosine in the quadruplex geometry. The geometry of the system restrained in the quadruplex form can lead to repulsive interactions with the other 2'-deoxyguanosines, so that it should be possible to explain the observation of blue shifts when solvent molecules are added.

Semiclassical spectra

So far we have considered only a classical nuclear framework to sample the Born-Oppenheimer surface of the system. In this framework the energy of the trajectory is ideally fixed during the dynamics and only the accessible configurations of the phase space will be sampled. We have employed short-time MD simulations and the sampled phase space in some cases could be insufficient. Propagating longer in time the MD simulations seems to be an appealing and easy solution, but this kind of approach brings in additional issues since more and more energy will be transferred to low frequency modes and even rigid rotations due to numerical inaccuracies and energy leakage. This would lead to a less accurate spectrum. One solution we find out to be helpful is to run more trajectories with same energy, as shown previously. Another solution is to consider an approximate quantum mechanical evolution of the system, as in the MC-DC-SCIVR method.[59–61] In this case a Gaussian wavepacket is propagated in the phase space, meaning that even by means of a single trajectory a wider area with respect to a classical trajectory is sampled. Since the two low frequency modes are strongly harmonic and in good agreement with the expected results, we focus again our attention on the symmetric and asymmetric $C2'H_2$ stretches.

We start again our analysis by considering the 2'-deoxyguanosine molecule. For the gas phase calculation to partition the system into subspaces we use the average Hessian on all the Hessians along the trajectory.[60] For the water box system the average Hessian is calculated starting from fifteen Hessians uniformly distributed along the trajectory. For the gas phase system we employ a bidimensional sub-
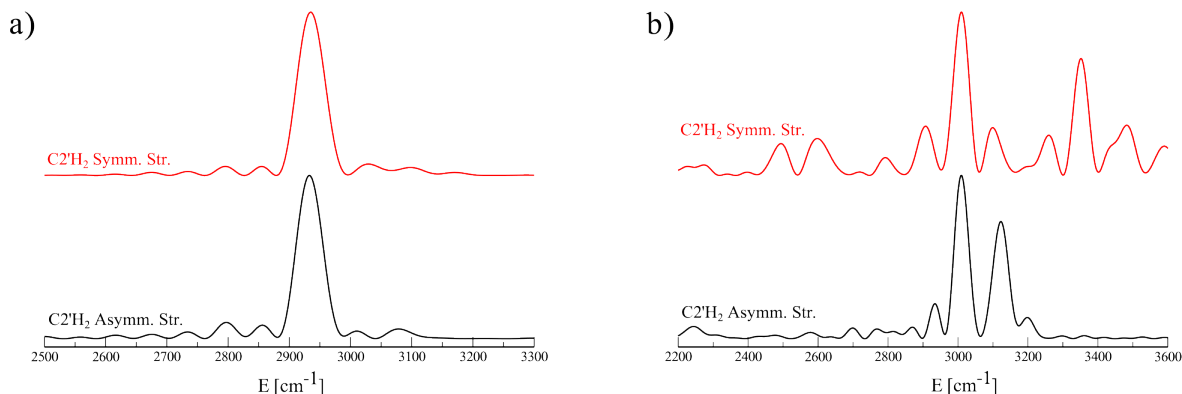
Figure 9: 2'-deoxyguanosine semiclassical spectra of the asymmmetric and symmetric stretching of the $C2'H_2$ group in gas (left) and water solution (right)

space in which the asymmetric and symmetric stretches are present. This means that these two normal modes are strongly coupled along the dynamics, and could also explain why only a large number of trajectories is able to reproduce a separation of the two signals.

The MC-DC SCIVR spectrum of 2'-deoxyguanosine in gas phase (figure 9 panel a) shows two signals that are nearly degenerate, which, similarly to the QCT case for the same molecule, is due to the fact that asymmetric/symmetric modes are strongly coupled to each other. Moving on to the water box system, we choose two monodimensional subspaces for both modes. By looking at the spectrum (figure 9 panel b) for the asymmetric mode a double peak is present, with the most intense peak still degenerate with the one of the symmetric mode (upper part of panel b). In this case by changing the initial conditions as done with 2'-deoxyguanosine, we were able to assign the less intense peak in fig. 9 b (black line) to the asymmetric stretch mode. We notice that by using MC-DC SCIVR to calculate the spectrum we obtain a better resolution with respect to the QCT counterpart. As far as the peak around 3350, we think that is a combination band originated by the $C2'H_2$ symmetric stretch.

Finally, we move to apply the semiclassical approach to the quadruplex employing the same protocol adopted for 2'-deoxyguanosine.

By analyzing the spectra in figure 10 we conclude that for the gas phase there is an improvement moving from QCT (figure s3 in the SI file) to SC single trajectory calculations. In particular, it is evident that most of the asymmetric and symmetric signals seem to be split. One problem that arise in this case
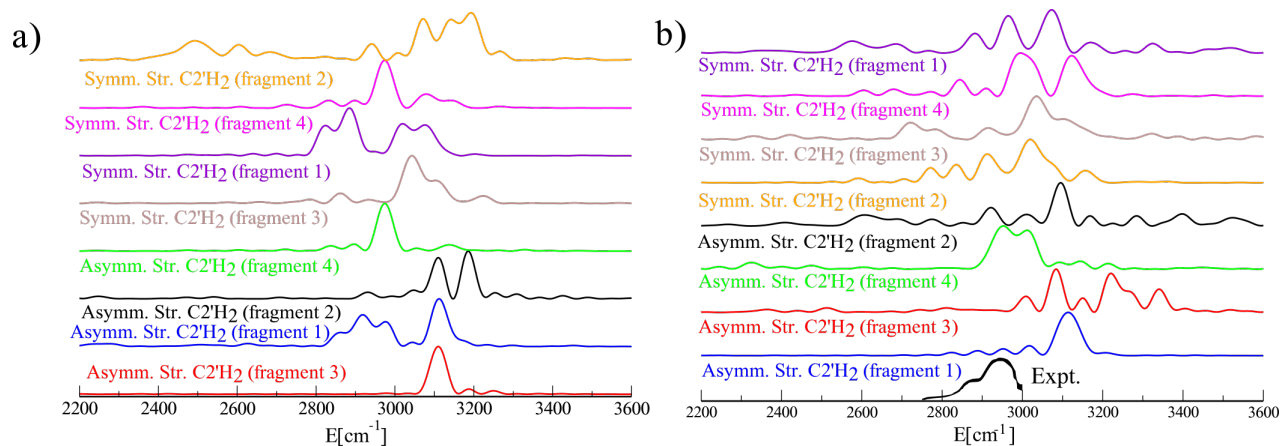
18

Figure 10: Semiclassical spectra of the quadruplex in gas phase a) and water solution b), for the asymmetric and symmetric stretches of $C2'H_2$.

is that the spectra are not always reporting a single or few peaks, due to the presence of resonances and other vibrational states strongly coupled to the $C2'H_2$ stretches, which hampers the assignment of the signals. The same is true also for the water box system, in which it is possible to observe that the experimental feature seems to be reproduced reasonably. On a quantitative level the experimental values in this case are about 100 cm^{-1} apart from the theoretical estimates, which seems to be fair since the employment of 100 trajectories gave nearly the same results for this set of modes.

Conclusions

With our analysis we were able to show how solvation acts differently on different vibrational features and compare them between the molecular and supramolecular arrangement. Solvation cannot be accounted for only by considering simple scaling factors, but it is often necessary to consider every normal mode and molecule explicitly employing a high level spectroscopy method. To corroborate our results we employed the radial pair function, which is able to assess how much a certain part of the molecule is able to interact with the solvent during the trajectory. FF do not provide highly precise descriptions of the PES and, for this reason, the use of FF in the theoretical spectroscopy field should be limited to a qualitative comparison rather than a quantitative one. In order to combine the possibility to deal with large size molecular systems and maintain an accurate description of the PES, QM/MM methods

19

should be employed. Relative to this matter a series of promising QM/MM methods are the ones that implement the mutual polarization of the QM and MM part.[76,77] In this work, we also showed that for some kind of normal modes a single trajectory with the "correct" initial conditions, can lead to a good qualitative agreement with respect to the experiment. Instead for other normal modes a distribution of trajectory run in the same energy shell is needed to increase the agreement with the experiment. By using FF it is possible to run a high number of trajectories without worrying too much about computational overheads. The last point in the conclusions, comes from the fact that we were also able to observe that a quantum mechanical description of the dynamics obtained by a single trajectory, is able to give us results similar to what it was obtained by calculating the spectra by means of multiple trajectories.

References

(1) Hoogsteen, K. *Acta Crystallographica* **1963**, *16*, 907–916.

(2) Kelly, R. E. A.; Lee, Y. J.; Kantorovich, L. N. *J. Phys. Chem. B* **2005**, *109*, 22045–22052.

(3) Sponer, J.; Mladek, A.; Spackova, N.; Cang, X.; Cheatham, T. E.; Grimme, S. Relative Stability of Different DNA Guanine Quadruplex Stem Topologies Derived Using Large-Scale Quantum-Chemical Computations. *J. Am. Chem. Soc.* **2013**, *135*, 9785–9796.

(4) Sieranski, T. *Theoretical Chemistry Accounts* **2020**, *139*, 153.

(5) Lipps, H. J.; Rhodes, D. G-quadruplex structures: in vivo evidence and function. *Trends in Cell Biology* **2009**, *19*, 414–422.

(6) Rhodes, D.; Lipps, H. J. G-quadruplexes and their regulatory roles in biology. *Nucleic acids research* **2015**, *43*, 8627–8637.

(7) Guiblet, W. M.; DeGiorgio, M.; Cheng, X.; Chiaromonte, F.; Eckert, K. A.; Huang, Y.-F.; Makova, K. D. Selection and thermostability suggest G-quadruplexes are novel functional elements of the human genome. *Genome Research* **2021**, *31*, 1136–1149.

(8) Adam, S.-J.; Grand Cory, L.; Bearss David, J.; Hurley Laurence, H. Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proceedings of the National Academy of Sciences* **2002**, *99*, 11593–11598.

(9) Rankin, S.; Reszka, A. P.; Huppert, J.; Zloh, M.; Parkinson, G. N.; Todd, A. K.; Ladame, S.; Balasubramanian, S.; Neidle, S. Putative DNA Quadruplex Formation within the Human c-kit Oncogene. *J. Am. Chem. Soc.* **2005**, *127*, 10584–10589.

(10) Cogoi, S.; Xodo, L. E. G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. *Nucleic Acids Res* **2006**, *34*, 2536–2549.

(11) Hansel-Hertsch, R.; Di Antonio, M.; Balasubramanian, S. DNA G-quadruplexes in the human genome: detection, functions and therapeutic potential. *Nature Reviews Molecular Cell Biology* **2017**, *18*, 279–284.

(12) Mohanty, B. K.; Karam, J. A. Q.; Howley, B. V.; Dalton, A. C.; Grelet, S.; Dincman, T.; Streitfeld, W. S.; Yoon, J.-H.; Balakrishnan, L.; Chazin, W. J.; Long, D. T.; Howe, P. H. Heterogeneous nuclear ribonucleoprotein E1 binds polycytosine DNA and monitors genome integrity. *Life Sci. Alliance* **2021**, *4*, e202000995.

(13) Musumeci, D.; Riccardi, C.; Montesarchio, D. G-Quadruplex Forming Oligonucleotides as Anti-HIV Agents. 2015.

(14) Perrone, R.; Nadai, M.; Frasson, I.; Poe, J. A.; Butovskaya, E.; Smithgall, T. E.; Palumbo, M.; Palu, G.; Richter, S. N. A Dynamic G-Quadruplex Region Regulates the HIV-1 Long Terminal Repeat Promoter. *J. Med. Chem.* **2013**, *56*, 6521–6530.

(15) Ruggiero, E.; Zanin, I.; Terreri, M.; Richter, S. N. G-Quadruplex Targeting in the Fight against Viruses: An Update. 2021.

(16) Liu, G.; Du, W.; Sang, X.; Tong, Q.; Wang, Y.; Chen, G.; Yuan, Y.; Jiang, L.; Cheng, W.; Liu, D.; Tian, Y.; Fu, X. RNA G-quadruplex in TMPRSS2 reduces SARS-CoV-2 infection. *Nature Communications* **2022**, *13*, 1444.

(17) Neidle, S. Quadruplex Nucleic Acids as Novel Therapeutic Targets. *J. Med. Chem.* **2016**, *59*, 5987–6011.

(18) Mergny, J.-L.; Phan, A.-T.; Lacroix, L. Following G-quartet formation by UV-spectroscopy. *FEBS Letters* **1998**, *435*, 74–78.

(19) Paramasivan, S.; Rujan, I.; Bolton, P. H. Circular dichroism of quadruplex DNAs: Applications to structure, cation effects and ligand binding. *Quadruplex DNA* **2007**, *43*, 324–331.

(20) Pagba, C. V.; Lane, S. M.; Wachsmann-Hogiu, S. *J. Raman Spectrosc.* **2010**, *41*, 241–247.

(21) Friedman, S. J.; Terentis, A. C. Analysis of G-quadruplex conformations using Raman and polarized Raman spectroscopy. *J. Raman Spectrosc.* **2016**, *47*, 259–268.

(22) Di Fonzo, S.; Amato, J.; D'Aria, F.; Caterino, M.; D'Amico, F.; Gessini, A.; Brady, J. W.; Cesaro, A.; Pagano, B.; Giancola, C. Ligand binding to G-quadruplex DNA: new insights from ultraviolet resonance Raman spectroscopy. *Phys. Chem. Chem. Phys.* **2020**, *22*, 8128–8140.

(23) Ivanov, A. Y. Conformations of ribonucleoside uridine in the low temperature Ar matrices. *Low Temperature Physics* **2010**, *36*, 458–464.

(24) Ivanov, A. Y.; Rubin, Y. V.; Egupov, S. A.; Belous, L. F.; Karachevtsev, V. A. The conformational structure of adenosine molecules, isolated in low-temperature Ar matrices. *Low Temperature Physics* **2015**, *41*, 936–941.

(25) Ivanov, A. Y.; Stepanian, S. G.; Karachevtsev, V. A.; Adamowicz, L. Nucleoside conformers in low-temperature argon matrices: Fourier transform IR spectroscopy of isolated thymidine and deuterothymidine molecules and quantum-mechanical calculations. *Low Temperature Physics* **2019**, *45*, 1008–1017.

(26) Asami, H.; Urashima, S.-h.; Saigusa, H. Hydration structures of 2-deoxyguanosine studied by IR-UV double resonance spectroscopy: comparison with guanosine. *Phys. Chem. Chem. Phys.* **2009**, *11*, 10466–10472.

(27) Choi, M. Y.; Miller, R. E. Four Tautomers of Isolated Guanine from Infrared Laser Spectroscopy in Helium Nanodroplets. *J. Am. Chem. Soc.* **2006**, *128*, 7320–7328.

(28) Oh, H.-B.; Lin, C.; Hwang, H. Y.; Zhai, H.; Breuker, K.; Zabrouskov, V.; Carpenter, B. K.; McLafferty, F. W. Infrared Photodissociation Spectroscopy of Electrosprayed Ions in a Fourier Transform Mass Spectrometer. *J. Am. Chem. Soc.* **2005**, *127*, 4076–4083.

(29) Wu, R.; McMahon, T. B. Infrared Multiple Photon Dissociation Spectra of Proline and Glycine

Proton-Bound Homodimers. Evidence for Zwitterionic Structure. *J. Am. Chem. Soc.* **2007**, *129*, 4864–4865.

(30) Fraschetti, C.; Montagna, M.; Guarcini, L.; Guidoni, L.; Filippi, A. Spectroscopic evidence for a gas-phase librating G-quartet-Na complex. *Chem. Commun.* **2014**, *50*, 14767–14770.

(31) Xu, M.; Gao, Y.; Moreno, E. M.; Kunst, M.; Muhler, M.; Wang, Y.; Idriss, H.; Woll, C. Photocatalytic Activity of Bulk Anatase and Rutile Single Crystals Using Infrared Absorption Spectroscopy. *Phys. Rev. Lett.* **2011**, *106*, 138302.

(32) Setvin, M. et al. A Multitechnique Study of CO Adsorption on the TiO2 Anatase (101) Surface. *J. Phys. Chem. C* **2015**, *119*, 21044–21052.

(33) Mondragon-Sanchez, J. A.; Liquier, J.; Shafer, R. H.; Taillandier, E. Tetraplex Structure Formation in the Thrombin-Binding DNA Aptamer by Metal Cations Measured by Vibrational Spectroscopy. *null* **2004**, *22*, 365–373.

(34) Gualerzi, A.; Picciolini, S.; Carlomagno, C.; Rodà, F.; Bedoni, M. Biophotonics for diagnostic detection of extracellular vesicles. *Advanced Drug Delivery Reviews* **2021**, *174*, 229–249.

(35) Conte, R.; Ceotto, M. *Semiclassical Molecular Dynamics for Spectroscopic Calculations*; Wiley Online Books; 2020; pp 595–628.

(36) Gabas, F.; Conte, R.; Ceotto, M. Semiclassical Vibrational Spectroscopy of Biological Molecules Using Force Fields. *J. Chem. Theory Comput.* **2020**, *16*, 3476–3485.

(37) Gabas, F.; Conte, R.; Ceotto, M. Quantum Vibrational Spectroscopy of Explicitly Solvated Thymidine in Semiclassical Approximation. *J. Phys. Chem. Lett.* **2022**, *13*, 1350–1355.

(38) Zhang, C.; Lu, C.; Jing, Z.; Wu, C.; Piquemal, J.-P.; Ponder, J. W.; Ren, P. *J. Chem. Theory Comput.* **2018**, *14*, 2084–2108.

(39) Rackers, J.; Wang, Z.; Lu, C.; Laury, M. L.; Lagardere, L.; Schnieders, M. J.; Piquemal, J.-P.; Ren, P.; Ponder, J. W. *J. Chem. Theory Comput.* **2018**, *14*, 5273–5289.

(40) Beeman, D. *Journal of Computational Physics* **1976**, *20*, 130–139.

(41) Miller, W. H. *Classical-limit quantum mechanics and the theory of molecular collisions*; Advances in chemical physics; 1974; Vol. 25; pp 69–177.

(42) Miller, W. H. The Semiclassical Initial Value Representation: A Potentially Practical Way for Adding Quantum Effects to Classical Molecular Dynamics Simulations. *The Journal of Physical Chemistry A* **2001**, *105*, 2942–2955.

(43) Miller, W. H. Quantum dynamics of complex molecular systems. *Proceedings of the National Academy of Sciences* **2005**, *102*, 6660–6664.

(44) Kay, K. G. The Herman-Kluk approximation: Derivation and semiclassical corrections. *Real-time dynamics in complex quantum systems* **2006**, *322*, 3–12.

(45) Heller, E. J. Frozen Gaussians: A very simple semiclassical approximation. *J. Chem. Phys.* **1981**, *75*, 2923–2931.

(46) Pollak, E. *Quantum Dynamics of Complex Molecular Systems*; Springer Berlin Heidelberg: Berlin, Heidelberg, 2007; pp 259–271.

(47) Grossmann, F. A semiclassical hybrid approach to many particle quantum dynamics. *J. Chem. Phys.* **2006**, *125*, 014111.

(48) Bonnet, L. Semiclassical initial value representation: From Møller to Miller. *The Journal of Chemical Physics* **2020**, *153*, 174102.

(49) Larregaray, P.; Bonnet, L. Including tunneling into the classical cross sections and rate constants for the $N(2D) + H2(v = 0, j = 0)$ reaction. *Theoretical Chemistry Accounts* **2021**, *140*, 1–7.

(50) Bonnet, L. Semiclassical descriptions of rotational transitions in natural and shifted angles: Analysis of unexpected results. *The Journal of Chemical Physics* **2021**, *155*, 174103.

(51) Begusic, T.; Tapavicza, E.; Vanicek, J. Applicability of the thawed Gaussian wavepacket dynamics to the calculation of vibronic spectra of molecules with double-well potential energy surfaces. *Journal of Chemical Theory and Computation* **2022**, *18*, 3065–3074.

(52) Begusic, T.; Vanicek, J. Finite-temperature, anharmonicity, and duschinsky effects on the two-dimensional electronic spectra from ab initio thermo-field Gaussian wavepacket dynamics. *The journal of physical chemistry letters* **2021**, *12*, 2997–3005.

(53) Vanicek, J.; Begusic, T. Ab initio semiclassical evaluation of vibrationally resolved electronic spectra with thawed Gaussians. *Molecular Spectroscopy and Quantum Dynamics* **2021**, 199–229.

(54) Begusic, T.; Vanicek, J. On-the-fly ab initio semiclassical evaluation of third-order response functions for two-dimensional electronic spectroscopy. *The Journal of Chemical Physics* **2020**, *153*, 184110.

(55) Malpathak, S.; Church, M. S.; Ananth, N. A Semiclassical Framework for Mixed Quantum Classical Dynamics. *The Journal of Physical Chemistry A*

(56) Church, M. S.; Ananth, N. Semiclassical dynamics in the mixed quantum-classical limit. *The Journal of chemical physics* **2019**, *151*, 134109.

(57) Church, M. S.; Hele, T. J.; Ezra, G. S.; Ananth, N. Nonadiabatic semiclassical dynamics in the mixed quantum-classical initial value representation. *The Journal of chemical physics* **2018**, *148*, 102326.

(58) Church, M. S.; Antipov, S. V.; Ananth, N. Validating and implementing modified Filinov phase filtration in semiclassical dynamics. *The Journal of chemical physics* **2017**, *146*, 234104.

(59) Ceotto, M.; Atahan, S.; Tantardini, G. F.; Aspuru-Guzik, A. Multiple coherent states for first-principles semiclassical initial value representation molecular dynamics. *J. Chem. Phys.* **2009**, *130*, 234113.

(60) Ceotto, M.; Di Liberto, G.; Conte, R. Semiclassical "Divide-and-Conquer" Method for Spectroscopic Calculations of High Dimensional Molecular Systems. *Phys. Rev. Lett.* **2017**, *119*, 010401.

(61) Di Liberto, G.; Conte, R.; Ceotto, M. "Divide and conquer" semiclassical molecular dynamics: A practical method for spectroscopic calculations of high dimensional molecular systems. *J. Chem. Phys.* **2018**, *148*, 014307.

(62) Gabas, F.; Di Liberto, G.; Conte, R.; Ceotto, M. Protonated glycine supramolecular systems: the need for quantum dynamics. *Chemical science* **2018**, *9*, 7894–7901.

(63) Bertaina, G.; Di Liberto, G.; Ceotto, M. Reduced rovibrational coupling Cartesian dynamics for semiclassical calculations: Application to the spectrum of the Zundel cation. *The Journal of Chemical Physics* **2019**, *151*, 114307.

(64) Aieta, C.; Micciarelli, M.; Bertaina, G.; Ceotto, M. Anharmonic quantum nuclear densities from full dimensional vibrational eigenfunctions with application to protonated glycine. *Nature communications* **2020**, *11*, 1–9.

(65) Gandolfi, M.; Rognoni, A.; Aieta, C.; Conte, R.; Ceotto, M. Machine learning for vibrational spectroscopy via divide-and-conquer semiclassical initial value representation molecular dynamics with application to N-methylacetamide. *The Journal of Chemical Physics* **2020**, *153*, 204104.

(66) Gabas, F.; Di Liberto, G.; Ceotto, M. Vibrational investigation of nucleobases by means of divide and conquer semiclassical dynamics. *The Journal of Chemical Physics* **2019**, *150*, 224107.

(67) Ceotto, M.; Zhuang, Y.; Hase, W. L. Accelerated direct semiclassical molecular dynamics using a compact finite difference Hessian scheme. *J. Chem. Phys.* **2013**, *138*, 054116.

(68) Conte, R.; Gabas, F.; Botti, G.; Zhuang, Y.; Ceotto, M. Semiclassical vibrational spectroscopy with Hessian databases. *J. Chem. Phys.* **2019**, *150*, 244118.

(69) Gandolfi, M.; Ceotto, M. Unsupervised Machine Learning Neural Gas Algorithm for Accurate

Evaluations of the Hessian Matrix in Molecular Dynamics. *J. Chem. Theory Comput.* **2021**, *17*, 6733–6746.

(70) Cazzaniga, M.; Micciarelli, M.; Moriggi, F.; Mahmoud, A.; Gabas, F.; Ceotto, M. Anharmonic calculations of vibrational spectra for molecular adsorbates: A divide-and-conquer semiclassical molecular dynamics approach. *J. Chem. Phys.* **2020**, *152*, 104104.

(71) Cazzaniga, M.; Micciarelli, M.; Gabas, F.; Finocchi, F.; Ceotto, M. Quantum Anharmonic Calculations of Vibrational Spectra for Water Adsorbed on Titania Anatase (101) Surface: Dissociative versus Molecular Adsorption. *The Journal of Physical Chemistry C* **2022**, *126*, 12060–12073.

(72) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N*log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(73) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.

(74) Levine, B. G.; Stone, J. E.; Kohlmeyer, A. Fast analysis of molecular dynamics trajectories with graphics processing units–Radial distribution function histogramming. *Journal of Computational Physics* **2011**, *230*, 3556–3569.

(75) Krafft, C.; Benevides, J. M.; Thomas, J., George J. Secondary structure polymorphism in Oxytricha nova telomeric DNA. *Nucleic Acids Res* **2002**, *30*, 3981–3991.

(76) Lipparini, F.; Mennucci, B. Hybrid QM/classical models: Methodological advances and new applications. *Chem. Phys. Rev.* **2021**, *2*, 041303.

(77) Macaluso, V.; Hashem, S.; Nottoli, M.; Lipparini, F.; Cupellini, L.; Mennucci, B. Ultrafast Transient Infrared Spectroscopy of Photoreceptors with Polarizable QM/MM Dynamics. *J. Phys. Chem. B* **2021**, *125*, 10282–10292.