Investigating the Spectroscopy of the Gas Phase Guanine-Cytosine Pair: Keto *vs* Enol Configurations

Giacomo Botti, Michele Ceotto, and Riccardo Conte*

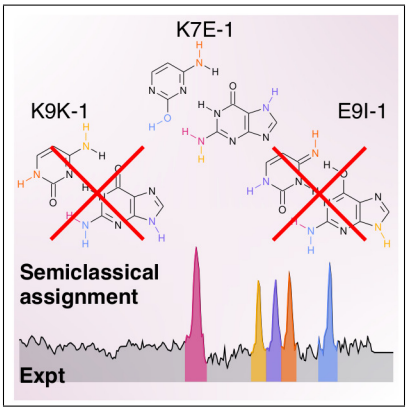
Dipartimento di Chimica, Università degli Studi di Milano, via Golgi 19, 20133 Milano, Italy

E-mail: riccardo.conte1@unimi.it

Abstract

We report on a vibrational study of the guanine-cytosine dimer tautomers using state-of-the-art quasi-classical trajectory and semiclassical vibrational spectroscopy. The latter includes possible quantum mechanical effects. Through an accurate comparison with the experimental spectra, we are able to shine a light on the hydrogen bond network of one of the main subunits of DNA and put on a solid footing the experimental assignment. Our calculations corroborate the experimental conclusion that the global minimum Watson-and-Crick structure is not detected in the spectra and there is no evidence of tunnel-effect-based double proton hopping. Our accurate assignment of the spectral features may also serve as a basis for the development of precise force fields to study the guanine-cytosine dimer.

TOC Graphic



Deoxyribonucleic acid (DNA) is a very important biopolymer because it stores the cell genetic information. Its duplication process heavily relies on the complementarity of nucleobases in forming hydrogen bonds. On the one hand, the nature of the hydrogen bond allows the DNA molecule to open and close the double helix, so that the genetic information can be duplicated by means of the DNA polymerase. On the other hand, this flexibility allows for the formation of other tertiary structures, such as B-DNA, A-DNA or even triple-stranded DNA.[1] Different base clusters are also possible, such as the Hoogsteen pairs and G-quadruplexes.[2,3] Given the presence of keto and amino groups, nucleobases present a great number of tautomers, a characteristic which can induce mismatches during the replication processes.[4] Even if ribonucleic acid (RNA) is the nucleic acid most affected by non-canonical pairing,[5] the nucleobase pair with the highest number of stable tautomers is guanine-cytosine (GC), which is in common with DNA. Therefore, we will focus on this pair.

One of the proposed mutagenic mechanisms is initiated by a proton hopping event, in which the two hydrogen atoms involved in the supramolecular bonding hop from one base to the other. As a consequence, new tautomers are generated and these are then mistakenly paired during DNA reproduction. However, the debate around the mechanism of this proton hopping is still open.[6–11] On the one hand, the double proton hopping process is thermally improbable and products are too short-lived to have any biological impact, even if this aspect does not exclude that the double proton hopping could happen as a defense mechanism, in a deactivation pathway.[10,12,13] On the other hand, the process could occur *via* tunneling, and this would be fast enough to make the product short life irrelevant.[8] If this is the case, the hopping tautomer could enter the polymerase λ pocket and be incorporated erroneously.[7] The discussion about tunneling in DNA bases is still open[9,14] and, at moment, this hypothesis has not been confirmed. More specifically, at the moment of our writing, no clear evidence of double proton hopping has been found, but the continuous development of fast and ultra-fast techniques might provide the experimentalists with the needed tools to catch this phenomenon if really present.

To unravel the DNA structural and mechanical peculiarities, it is necessary to understand the equilibrium and — more importantly — the dynamical properties of the pairs of nucleobases, and focus on the hydrogen bond network. We think that understanding the dynamics of the hydrogen couple confined between the two nucleobases is pivotal to develop an accurate model for *in vivo* DNA stability, and also for base tautomerization and mismatch, even more if quantum effects are to be included.

To this aim we employ vibrational spectroscopy because the effects of hydrogen bonding can be easily detected through it, and non-trivial information can be extracted when the vibrational spectra of hydrogen bonded species are compared with the corresponding non-bonded ones.[15–28] Therefore, the experimental vibrational spectra of nucleobase dimers are interesting and well-known in gas phase, where the hydrogen bond contribution to the pair formation and stability can be isolated from the π-stacking and hydrophobic one. The hydrogen-bonded structure is observed only in gas phase, since in aqueous solution the bases prefer the stacked disposition.[29,30] The vibrational spectra could also help to clarify if any proton hopping is taking place, in particular in the guanine-cytosine pair, by inspecting if specific features of the proton-hopped tautomer are present in the spectra.[29,31–36] The main current limitation of these experiments is the absence of a solid and thorough assignment of the main spectral features, which would allow one to clearly identify the corresponding tautomer. Specifically, the experimental spectrum of the GC dimer is currently assigned to a high energy tautomer (named K7E-1), instead of the expected global minimum one, which is the Watson and Crick (WC, or K9K-1) structure.[29,37] This spectrum has been obtained by IR-UV hole burning spectroscopy on a mixture of laser-desorbed guanine and cytosine.[32,33]

The interest in comparing our simulations to the experimental findings is twofold. The first goal is to check if the WC structure is really not present in the experimental spectrum. Nir et al. based their assignment on a scaled-harmonic approach, which is often a non-reliable and *ad hoc* technique not able to include quantum effects, and on the isolated bases computational setup spectra, which do not account for inter-molecular effects. Therefore, it

is possible that a precise quantum assignment of the WC tautomer spectrum matches the experimental spectrum demonstrating the presence of relevant quantum effects. However, if conversely the absence of the WC tautomer in the spectrum is confirmed, then this fact would support the hypothesis that the WC excited state lifetime is unexpectedly short.[29,32,38] Then, we want to check out if the vibrational spectrum may reveal the fingerprints of quantum effects related to the double proton hopping mechanism. There are two main ways whereby we can point out this phenomenon: We can simulate the molecular species which would be the result of the double proton hopping starting from the WC tautomer, i.e. an enolic form labeled as E9K-1, and check if it is present in the experimental spectrum; furthermore, it is possible to look for differences in the spectral features between calculated WC tautomer spectra obtained by means of a theoretical method able to point out quantum effects and another method unable to do that. In general an accurate assignment of the experimental spectrum by means of a refined theoretical technique will be of great help to build force field models to be employed in the study of the GC dimer.

Current computational and theoretical vibrational studies of DNA bases are limited to static (harmonic-like) approximations or low-dimensional models, due to the system intrinsic complexity.[32,34,39,40] These approximations fall short when dealing with hydrogen bonded systems,[28] and a higher accuracy is necessary if one wants to assign nucleobase pair experimental spectra, where there are a pletora of tautomers.[31–33] To overcome these limitations, we employ the Divide-and-Conquer Semiclassical Initial Value Representation (DC SCIVR) method to compute the vibrational power spectra of the guanine-cytosine dimer.[41] DC SCIVR is an acknowledged method, capable of accounting for anharmonicity and reproducing quantum effects, such as zero-point energy (zpe), overtones and combination bands, using a single classical trajectory.[42–44] This allows us to limit the computational effort, even when investigating the 29-atom guanine-cytosine pair. DC SCIVR has already been applied with success to the vibrational study of isolated and solvated nucleobases, and to other nucleotide-based macromolecules.[3,24,45,46] The goal of these DC-SCIVR spectra is to assign

the experimental features on a solid footing and clear the open issues about both structure and stability of the nucleobase pairs.

The Time-Averaged Semiclassical Initial Value Representation (TA SCIVR)[47–50] is a quantum approximate method which can be applied to spectroscopy calculations of moderate dimension systems. In TA SCIVR, the vibrational power spectrum $I(E)$ can be computed as

$$I(E) = \left(\frac{1}{2\pi\hbar}\right)^{N_v} \iint d\mathbf{p}_0 d\mathbf{q}_0 \, \frac{1}{2\pi\hbar T} \left| \int_0^T dt \, e^{\frac{i}{\hbar}[S_t(\mathbf{p}_0,\mathbf{q}_0)+Et+\phi_t(\mathbf{p}_0,\mathbf{q}_0)]} \langle \Psi | g_t(\mathbf{p}_0,\mathbf{q}_0)\rangle \right|^2, \quad (1)$$

where E is the vibrational energy, N_v is the number of vibrational degrees of freedom, $(\mathbf{p}_0, \mathbf{q}_0)$ are the starting conditions, T is the total simulation time, $S_t(\mathbf{p}_0, \mathbf{q}_0)$ and $\phi_t(\mathbf{p}_0, \mathbf{q}_0)$ are the instantaneous classical action and the phase of the Herman-Kluk pre-exponential factor[51] respectively, and $\langle \Psi | g_t(\mathbf{p}_0, \mathbf{q}_0)\rangle$ is the quantum overlap between an arbitrary reference state $|\Psi\rangle$ and a coherent state evolved for a time t ($|g_t(\mathbf{p}_0, \mathbf{q}_0)\rangle$). Additional details can be found in the Supporting Information.

The Multiple Coherent Semiclassical Initial Value Representation was introduced to adapt the TA-SCIVR method to *on-the-fly* calculations.[52–55] MC SCIVR is based on two pillars: first, a single and tailored semiclassical trajectory at the exact quantum energy is able to fully describe the quantum state;[56] secondly, it is possible to increase the vibrational signal collected by this single trajectory by defining the reference state $|\Psi\rangle$ by means of an appropriate combination of two coherent states. Therefore, when employing MC SCIVR, a single tailored trajectory is enough to compute an accurate vibrational power spectrum.[55]

However, when dealing with large systems, the curse of dimensionality sets in. This is a decrease in the signal to noise ratio of the power spectrum, caused by the superposition integral in Eq. (1), which goes to zero when the number of degrees of freedom increases. To tackle this curse, the Divide and Conquer (DC) technique was developed.[41,57] In DC SCIVR the semiclassical power spectrum is computed on a subspace of the full dimensional

space while the trajectory is still evolved in full dimensionality and partly able to recollect interactions between modes belonging to different subspaces. Additional details on these semiclassical approaches, useful to help to replicate results, can be found in the Supplementary Information.

When performing a semiclassical calculation, it can be useful to compute also the classical velocity autocorrelation spectra employing the same MC-SCIVR classical trajectory as a term of comparison, thus performing a quasi-classical Trajectory (QCT) calculation. The QCT spectrum of the j-th vibrational normal mode is evaluated as[58,59]

$$I_j(E) = \frac{1}{2T} \left| \int_0^T e^{iEt/\hbar} p_j(t)\, dt \right|^2 , \tag{2}$$

where T is the total classical trajectory time, E is the energy, and $p_j(t)$ is the linear momentum of the j-th vibrational normal mode at time t. We perform *ab initio* "on-the-fly" Cartesian evolution of the dynamics with Cartesian coordinates and momenta transformed into normal mode coordinates and momenta at each time step along the trajectory. The QCT spectrum is able to include the potential energy surface anharmonicity, but it cannot collect any quantum effect. Nonetheless, it provides affordable and important data, and it has been applied with success to several systems.[15,28,58]

The conformational landscape of the GC pair is quite variegated because it depends on the tautomeric forms of both guanine and cytosine and their relative positions, *i.e.* the hydrogen bond network. Therefore, we adopt the nomenclature established in the experimental literature.[32,33] According to this nomenclature, the GC pair tautomer is identified by indicating: i) the guanine tautomer, which can be either enol- (E) or keto- (K); ii) the position of the imidazolic hydrogen of guanine, which is either on nitrogen 7 or 9; iii) the tautomer of cytosine which is keto- (K), enol- (E) or imino- (I); iv) a number related to the relative energy of the hydrogen-bond network. In this nomenclature, the Watson and Crick canonical tautomer is named K9K-1.

We start by studying the conformational landscape of the four tautomers lying lower in energy and reported in Fig.(1). K9K-1 is the canonical and most stable tautomer, i.e. the WC tautomer. Then, we consider the K7E-1 tautomer, because the literature suggests its presence in the experimental spectra. Furthermore, we focus on the K9E-1 tautomer since it is indistinguishable from K7E-1 when looking at the high harmonic frequencies, and it is the more similar to K9K-1. Finally, we consider also the E9I-1 tautomer because it is formed upon double proton hopping starting from K9K-1, and the presence of E9I-1 — even in traces — would suggest the presence of a fast-proceeding double proton hopping mechanism resulting in a stable tautomeric form.[7–10,12,32]

We optimize all the anticipated tautomers at DFT-D/B3LYP level of theory with def2-TZVP basis set,[60] where DFT-D stands for DFT with Grimme's empirical dispersions.[61] The equilibrium geometries with relative energies are shown in Fig. 1. The complete energy analysis is reported in Table S2 of the Supplementary Information (SI). The relative energies are in good agreement with those reported in the literature, confirming the WC tautomer (*i.e.* K9K-1) as the global minimum of the hydrogen-bonded structure of the GC pair. These four conformers present very different energies, even if the H bond structure is quite similar.

Moving to spectroscopy, our reference experiments are those collected by de Vries *et al.* using the hole burning IR-UV spectroscopy in a TOF spectrometer.[29] The experimental spectra were assigned using the spectra of the isolated bases and scaled harmonic frequencies at the RI-MP2 TZVPP *ab initio* level of theory.[29,32,33] The spectrum obtained by laser ablation of guanine and cytosine was initially assigned as either K7E-1 or K9E-1.[32,33] Upon investigation of the low frequency region, it was concluded that the compound responsible for the spectrum was K7E-1, given the position of the N^7H bending signal. The experimental assignment is summarized in the SI in Table S3.

In our simulations we start calculating the QCT power spectra of each tautomer, which we report in Fig. 2. QCT spectra are obtained by means of Eq. (2) runs based on a 25 000 au trajectory at DFT-D/B3LYP def2-TZVP level of theory. The focus is on modes above 3300 cm^{-1}

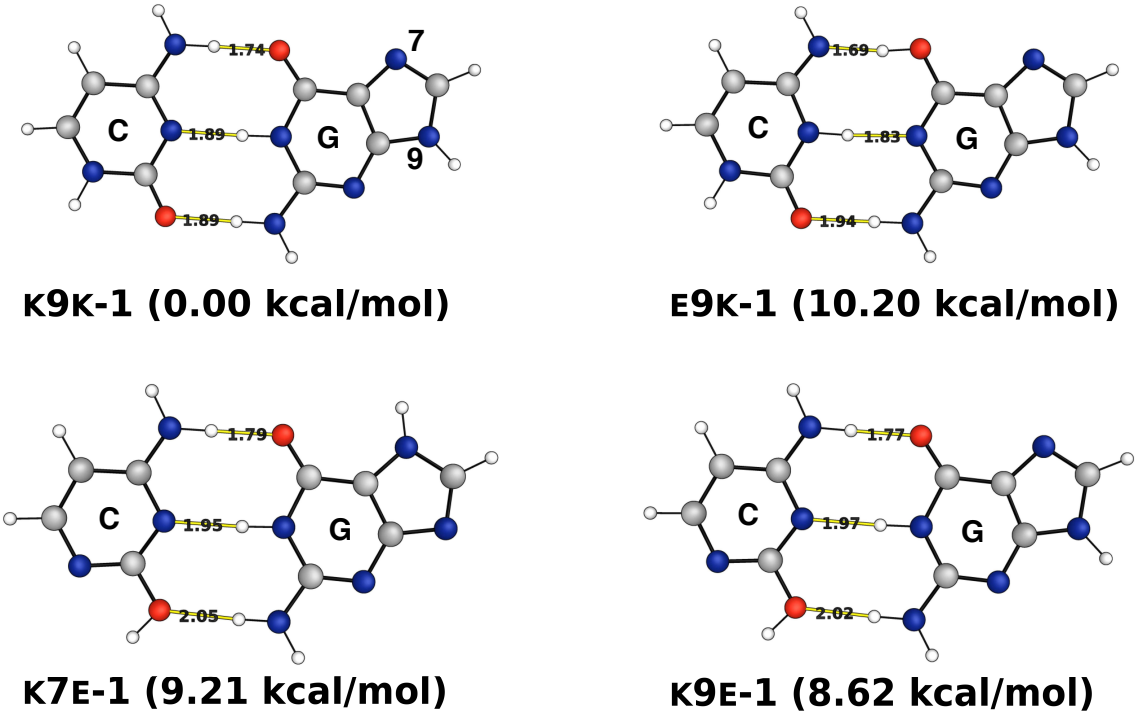


Figure 1: Equilibrium geometries and energies (in kcal mol^{-1}) of the four investigated GC tautomers GC tautomers, at DFT-D/B3LYP def2-TZVP level of theory. The hydrogen bonds are indicated by the yellow lines, with distances in Å.

for each tautomer. Furthermore, the QCT values are compared with the harmonic frequencies in Table S4 of the Supporting Information to appreciate the level of anharmonicity of the system. In the high frequency region, the QCT power spectra of K9K-1 and E9I-1 reported on the upper part of Fig. 2 are different, mainly in the guanine NH_2 stretch. This suggests that the main difference in this frequency region is due to the O···HNH hydrogen bond effects on the guanine NH_2 group. Given that QCT, in absence of relevant quantum effects, is generally a pretty accurate method, this feature provides a way to confirm or deny the presence of E9I-1 in K9K-1 gas-phase spectra. Conversely, in the lower part of Fig. 2, the QCT power spectra of K7E-1 and K9E-1 in the same high frequency region are basically indistinguishable from each other within the unavoidable peak width originated by the finite-time Fourier transform.

As a comparison with these QCT results, we employ DC SCIVR for K9K-1 and K7E-1. We do not consider for DC-SCIVR simulations the E9I-1 tautomer since it is populated only at high energies. Instead we still consider the K7E-1 tautomer because it is reported as the main tautomer in experimental spectra, even if K9K-1 and K9E-1 are lower in energy.[29,32–34,37] The DC-SCIVR spectra are shown in Fig. 2 as black solid lines and they confirm the QCT spectra values.

The only way the experimentalists had to obtain a vibrational spectrum related to K9K-1 was to lock the tautomer by alkylation, thus collecting the spectrum of *ethyl*-K9-*methyl*-K-1, so we begin the assignment from the *ethyl*-K9-*methyl*-K-1 spectrum.[32] We make the assumption that alkylation has no other effect in the high-frequency range than removing the corresponding NH stretches. The graphical comparison is shown in Fig. 3 and the analysis is reported in Table 1.

All the spectral features are assigned with good agreement between the experimental and computed frequencies. All the computed frequencies are blue shifted compared with the experiment. In particular, the guanine NH stretching in NH_2 (Gua NH@NH_2) is associated with the broad signal around 3280 cm^{-1}. The broad signal is an effect of the strong hydrogen

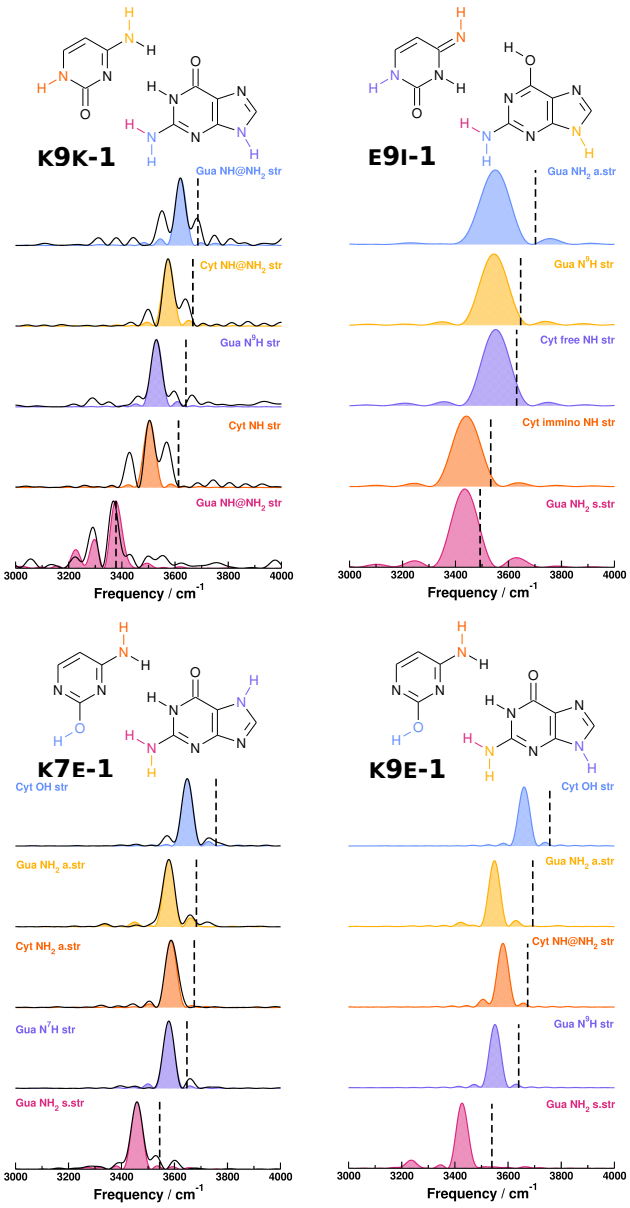


Figure 2: Power spectra in the frequency region above 3300 cm^{-1} for the four tautomers, obtained with QCT on a 25 000 au long trajectory, starting from harmonic conditions at DFT-D/B3LYP def2-TZVP level of theory. The harmonic frequencies are reported as black dashed lines. For K9K-1 and K7E-1 the DC-SCIVR spectra are reported in black solid lines. See Table S4 of the Supplementary Information (SI) and Table 1 for the numerical values

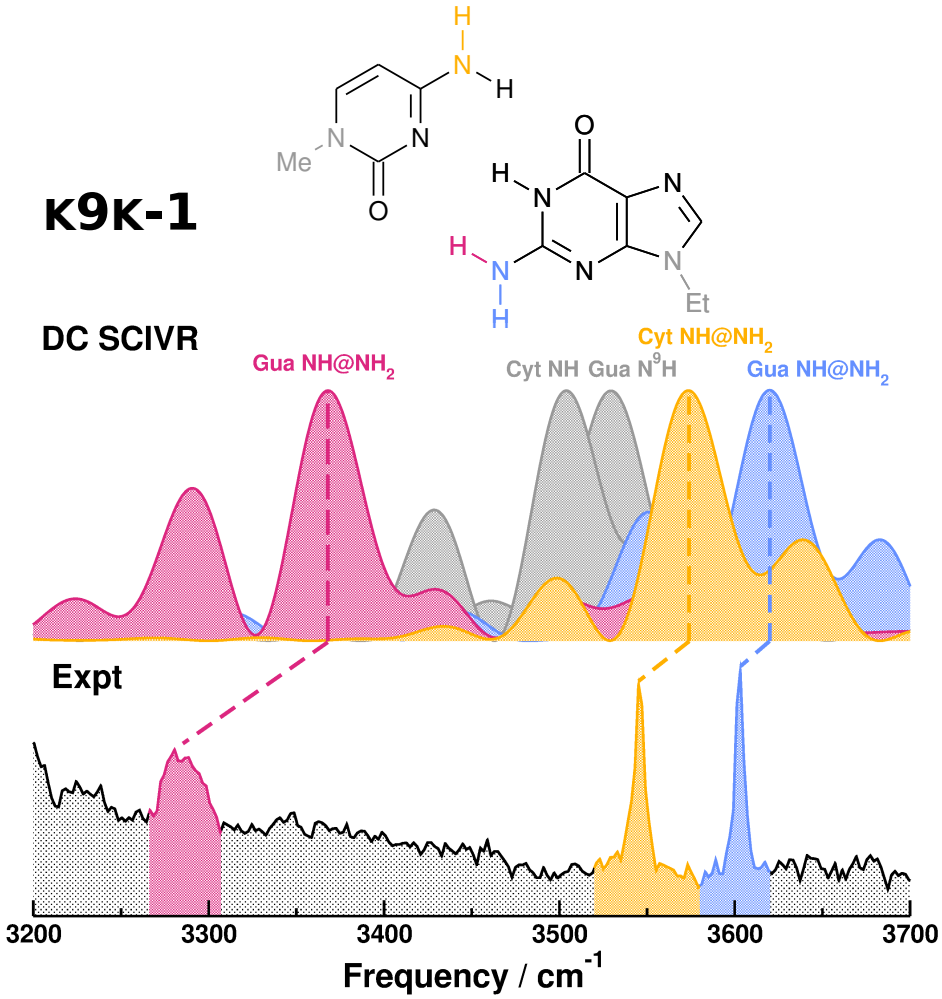


Figure 3: Comparison between the experimental spectra of *ethyl*-K9-*methyl*-K-1 and the DC-SCIVR spectra of K9K-1, obtained from a single 25 000 au trajectory, starting from harmonic conditions at DFT-D/B3LYP def2-TZVP level of theory. The grey peaks are the NH stretches suppressed by the experimental alkylation. The experimental spectrum is reproduced from ref. 29 (Copyright (2004) National Academy of Sciences, U.S.A.).

bond, while the larger disagreement between the experimental and DC-SCIVR frequency of this vibrational mode (compared to the other modes) is probably due to the chosen DFT-D/B3LYP level of theory. Indeed, this vibrational frequency is very similar for the harmonic approximation, QCT and DC-SCIVR estimates, even for different basis sets (see SI, Table S5). This is quite an unusual behavior and it suggests that either there is an improper description of the potential energy surface or our assumption breaks down and the experimental alkylation makes a difference for this mode. Nonetheless, we are able to assign this feature, since this is the only signal compatible with the guanine NH stretch in the NH_2 DC-SCIVR frequency. In conclusion, we can exclude the presence of the E9I-1-like tautomer, since there is no experimental signal that we can compare to E9I-1 guanine NH_2 stretches. This means that we find no spectral evidence of a double proton hopping mechanism.

We then proceed to the assignment of the experimental spectrum obtained by laser desorption of guanine and cytosine, starting with the K7E-1 DC-SCIVR spectra, as suggested in the literature.[29,32,34] The graphical comparison is shown in Fig. 4 and the analysis is reported in Table 1. Our assignment of the experimental signals differs from the one previously reported by de Vries *et al.*, and the main reason is that DC SCIVR fully accounts for mode anharmonicity while a rough scaled harmonic approximation has been previously employed by those same authors for their assignment. Specifically, the modes responsible for the three-pronged structure of the experimental spectrum are in different order than the previous experimental assignment: the peak at $3561\,cm^{-1}$ is assigned to cytosine NH_2 stretching instead of guanine NH_2 asymmetric stretching; the peak at $3543\,cm^{-1}$ is assigned to guanine N^7H stretching instead of cytosine NH_2 stretching; the peak at $3520\,cm^{-1}$ is assigned to guanine NH_2 asymmetric stretching instead of guanine N^7H stretching. Our main argument to assign the experimental spectrum to the K7E-1 tautomer and rule out the K9K-1 one is that the experimental spectrum lacks the broad signal at around $3280\,cm^{-1}$, which is present in the *ethyl*-K9-*methyl*-K-1 spectrum due to the K9K-1 guanine NH stretch in NH_2, a mode heavily influenced by the hydrogen bond with the cytosine ketonic function.

The experimental spectrum shown in Fig. 4 is instead characterized by a sharp peak at 3436 cm^{-1}, compatible with the K7E-1 guanine symmetric NH_2 stretch. We can also exclude the presence in the experiment of the e_9i-1 tautomer. First of all because e_9i-1 would be obtained by means of double proton hopping from the more stable k_9k-1 tautomer and if the latter is missing it is unlikely that the former is present. Then, because the highest frequency mode of e_9i-1, the cytosine NH stretch at 3552 cm^{-1}, is too low in frequency to match the experimental signal at 3615 cm^{-1}. This is in spite of our QCT and DC-SCIVR frequencies being even blue shifted compared to the experiment, presumably due to the approximate description of hydrogen bonds and other interactions at the chosen affordable level of electronic theory. It remains to rule out the presence of K9E-1, which is lower in energy than K7E-1 and whose high frequency spectrum is basically indistinguishable from that of K7E-1. To this end we performed two QCT simulations (one per tautomer) of the out-of-plane NH bending because de Vries and coworkers could not assign any signal above 500 cm^{-1} in their experiments to this vibrational mode. The scaled harmonic calculations suggested for the out-of-plane NH bending a frequency of 477 cm^{-1} for the K7E-1 tautomer, and a frequency of 508 cm^{-1} for the K9E-1 tautomer.[34] The missing peak just above 500 cm^{-1} allowed de Vries and coworkers to rule out the presence of K9E-1 in the experimental spectra. Our QCT simulations confirm and strengthen this conclusion estimating the target bending at 426 cm^{-1} for K7E-1 and at 509 cm^{-1} for K9E-1. Therefore, following de Vries' reasoning, we also rule out the presence of K9E-1 in the experimental spectra. A figure reporting the outcome of QCT calculations can be found in the Supplementary Information (see Figure S1).

Using DC-SCIVR and QCT simulations we obtained the vibrational power spectra of two tautomers of the guanine-cytosine pair. By comparing our results with the experimental spectra found in the literature,[29,32] we managed to assign the relevant spectral features in the high frequency region of two experimental spectra, one for the isolated guanine-cytosine dimer and the other one for its alkylated form. Indeed, the presence of a peak at 3436 cm^{-1} in

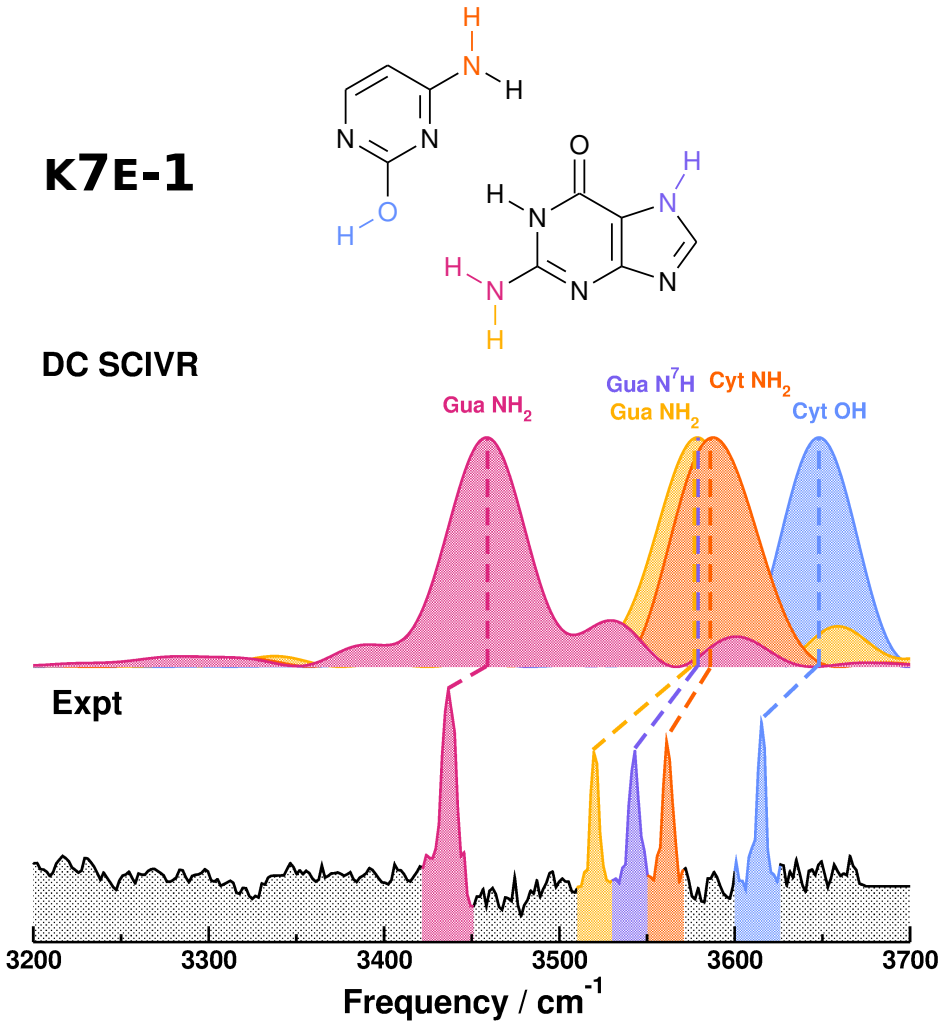


Figure 4: Comparison between the experimental spectra of the guanine-cytosine dimer and the DC-SCIVR spectra of K7E-1, obtained from a single 25 000 au trajectory, starting from harmonic conditions at DFT-D/B3LYP def2-TZVP level of theory. The experimental spectrum is reproduced from ref. 29 (Copyright (2004) National Academy of Sciences, U.S.A.).

Table 1: Comparison between the DC-SCIVR vibrational frequencies and the experimental ones (in cm^{-1}) for the main guanine-cytosine pair tautomers.

K9K-1			
Mode	DC SCIVR[(a)]	Scaled Harm.[(b)]	Expt.[(c)]
Gua NH str in NH_2	3620	3538	3603
Cyt NH str in NH_2	3574	3526	3545
Gua N^9H str	3529	3505	alkylated
Cyt NH str	3504	3476	alkylated
Gua NH str in NH_2	3368	3343	3283
MAE[(d)]	43	48	

K7E-1			
Mode	DC SCIVR	Scaled Harm.	Expt
Cyt OH str	3656	3595	3615
Gua NH_2 a.str	3534	3547	3520
Cyt NH_2 a.str	3559	3531	3561
Gua N^7H str	3536	3511	3543
Gua NH_2 s.str	3448	3419	3436
MAE[(e)]	15	25	

(a) The DC-SCIVR frequencies are obtained from a 25 000 au classical trajectory at DFT-D/B3LYP def2-TZVP level of theory for each tautomer.

(b) Scaled harmonic frequencies obtained at level of theory HF/6-31G(d,p), using scaling factor 0.893 for NH stretching and 0.867 for OH stretching.[32]

(c) The experimental frequencies are taken from the *ethyl*-K9-*methyl*-K-1 spectrum and the guanine-cytosine pair spectrum (K7E-1).[29]

(d) Mean Absolute Error (MAE) is calculated on only 3 data against experimental results.

(e) MAE is calculated on 5 data against experimental results.

the spectrum of the guanine-cytosine dimer is the signature feature of a high energy keto-enol tautomer (K7E-1), since the NH peak in the experimental spectrum of the alkylated Watson-and-Crick tautomer (K9K-1) is at lower frequency. This peak corresponds to the vibration of the guanine NH_2 group, which points towards either the ketonic (for K9K-1) or enolic (for K7E-1) form of cytosine. In K7E-1 the peak has been anticipated at higher frequencies ($3448\,cm^{-1}$), with a longer O···HN distance (2.05 Å), whereas K9K-1 shows the corresponding peak at lower frequencies ($3368\,cm^{-1}$) and a shorter O···HN distance (1.89 Å). In terms of normal modes, K7E-1 is characterized by the usual symmetric/asymmetric motions for the NH_2 stretching. Conversely, in K9K-1 the two NH of the guanine NH_2 are uncoupled and vibrate separately. This means that in K7E-1 the hydrogen bond is weaker than the one in K9K-1, resulting in a smaller perturbation of the original guanine free NH_2 vibrational motion. The fact that the corresponding E9I-1 mode has a QCT frequency of $3435\,cm^{-1}$ and a O···HN distance of 1.94 Å clarifies that the frequency of the guanine NH stretch does not depend only on the hydrogen bond acceptor, but also on the relative position of the two bases. Sometimes QCT and DC-SCIVR spectra present several peaks (see, for instance, Fig. 2). While it is evident which signal correspond to the target one, side peaks may be due to modes coupled to the target one, in the case of QCT simulations, or to additional combination bands/overtones not detected by means of QCT in the case of the DC-SCIVR calculations.

Our assignment agrees with the literature,[29,32,33,37] where the spectrum of Fig. (4) is assigned to the higher energy keto-enol form (K7E-1) and there is no fingerprint of the global minimum Watson-and-Crick form (WC — K9K-1). The WC form distinctive features appear in the experimental spectrum only when the keto-enol tautomerization is prevented by alkylation.[29] The absence of the WC form is probably due to the short lifetime of its excited state in the UV pump phase.[29,32] Theoretical calculations by Domcke *et al.* suggest that an internal conversion coadiuvated by the crossing with two doorway states may occur.[38]

It must be noted that in the high frequency region we cannot distinguish between K7E-1

and K9E-1, given that their QCT signals lie well within the peak width. However, this does not jeopardize our conclusions, since K7E-1 and K9E-1 only differ for the position of a hydrogen not involved in the hydrogen bonds. Furthermore, an examination of the low frequency region has allowed us to differentiate between the two tautomers and to rule out the presence of K9E-1 in the spectrum.

Eventually, we can rule out the presence of the tautomeric form (E9I-1) originated via double proton hopping from the Watson-and-Crick form (WC — K9K-1). Indeed, the distinctive signals of this enol-tautomer are absent in the experiments. Specifically, the highest frequency transition — attributed to the cytosine free NH stretching — is at too low frequency values and the guanine NH_2 asymmetric stretching is absent. This suggests that no hopping mechanism is detectable for the nucleobase pair under the IR-UV hole burning experiment conditions. In the past we demonstrated that NH_2 rotors can be characterized by quantum spectroscopic features, which can be detected by DC-SCIVR simulations but not by QCT ones.[25] This aspect together with the hypothesis that a quantum effect like tunneling could play a major role in the interconversion of the guanine-cytosine tautomers, led us to perform calculations employing both DC SCIVR and QCT. The result is that there are not clear differences in the spectra obtained with the two approaches, with equivalent fundamental frequencies of vibrations, strengthening our conclusions. Furthermore, in our calculations for the power spectra of E9I-1 we find that peaks are characterized by a much larger full width at half maximum (about $125\,\mathrm{cm}^{-1}$ against about $50\,\mathrm{cm}^{-1}$ for the other tautomers). This is an evidence that E9I-1 is a metastable state and it is about to convert to the WC tautomer in spite of the very short time of the simulation (about 600 fs). This is in agreement with recent calculations of the kinetic constant for the conversion from E9I-1 to K9K-1 by Angiolari et al.[62] Those calculations illustrate that this process is very fast and influenced by the environment. If the mutagenesis of base coupling is mediated by the double proton hopping mechanism it means that it happens in a very short time, much faster than conversion to the WC form. We notice that quantum effects are usually more evident

in the gas phase than for solvated systems, so our investigation hints at the possibility that quantum effects do not play a major role neither for DNA in solution. More investigations are indeed necessary on this point. For a better characterization of the mechanism, future spectroscopic investigations (both at experimental and theoretical level) should include the role of the environment.

Associated Content

Supporting Information

The Supporting Information are available free of charge at link. Contents: (1) Low-frequency analysis for K7E-1 and K9E-1 (2) Tautomer energies, summary of experimental assignment and comparison of the harmonic, comparison between harmonic and anharmonic (QCT) frequencies, QCT and DC SCIVR frequencies for K9K-1 NH stretching in NH_2; (3) further informations on the semiclassical methods employed

Additional Data

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

Author Informations

Notes

The authors have no conflict of interest to disclose.

Acknowledgments

The authors thank CINECA for the computational resources granted under the HPC project IsC99_SemiGC. The authors also thank Prof. F. Finocchi, Dr. E. Fallacara, Dr. S. Huppert and Dr. P. Depondt at Institut des NanoSciences de Paris (Sorbonne University) for useful discussions. G.B. thanks INPS for funding his Ph.D. scholarship. R.C. thanks Università degli Studi di Milano under grant PSR2021−DIP−005−RCONT. All authors thank the ERC for funding under project ERC POC SEMISOFT (Grant No. 101081361).

References

(1) Kim, S. K.; Takahashi, M.; Nordén, B. Binding of RecA to Anti-Parallel Poly(dA) · 2poly(dT) Triple Helix DNA. *Biochim. Biophys. Acta - Gene Structure and Expression* **1995**, *1264*, 129–133.

(2) Hoogsteen, K. The Crystal and Molecular Structure of a Hydrogen-Bonded Complex between 1-Methylthymine and 9-Methyladenine. *Acta Cryst* **1963**, *16*, 907–916.

(3) Moscato, D.; Gabas, F.; Conte, R.; Ceotto, M. Vibrational Spectroscopy Simulation of Solvation Effects on a G-quadruplex. *J. Biomol. Struct. Dyn.* **2023**, 1–11.

(4) Watson, J. D.; Crick, F. H. C. The Structure of DNA. *Cold Spring Harb. Symp. Quant. Biol.* **1953**, *18*, 123–131.

(5) Lemieux, S. RNA Canonical and Non-Canonical Base Pairing Types: A Recognition Method and Complete Repertoire. *Nucleic Acids Res.* **2002**, *30*, 4250–4263.

(6) Löwdin, P.-O. Proton Tunneling in DNA and Its Biological Implications. *Rev. Mod. Phys.* **1963**, *35*, 724–732.

(7) Slocombe, L.; Al-Khalili, J. S.; Sacchi, M. Quantum and Classical Effects in DNA Point

Mutations: Watson–Crick Tautomerism in AT and GC Base Pairs. *Phys. Chem. Chem. Phys.* **2021**, *23*, 4141–4150.

(8) Slocombe, L.; Sacchi, M.; Al-Khalili, J. An Open Quantum Systems Approach to Proton Tunnelling in DNA. *Commun. Phys.* **2022**, *5*, 109.

(9) Slocombe, L.; Winokan, M.; Al-Khalili, J.; Sacchi, M. Quantum Tunnelling Effects in the Guanine-Thymine Wobble Misincorporation via Tautomerism. *J. Phys. Chem. Lett.* **2023**, *14*, 9–15.

(10) Soler-Polo, D.; Mendieta-Moreno, J. I.; Trabada, D. G.; Mendieta, J.; Ortega, J. Proton Transfer in Guanine-Cytosine Base Pairs in B-DNA. *J. Chem. Theory Comput.* **2019**, *15*, 6984–6991.

(11) Umesaki, K.; Odai, K. A Kinetic Approach to Double Proton Transfer in Watson–Crick DNA Base Pairs. *J. Phys. Chem. B* **2020**, *124*, 1715–1722.

(12) Gheorghiu, A.; Coveney, P. V.; Arabi, A. A. The Influence of Base Pair Tautomerism on Single Point Mutations in Aqueous DNA. *Interface Focus* **2020**, *10*, 20190120.

(13) Hartweg, S.; Hochlaf, M.; Garcia, G. A.; Nahon, L. Photoionization Dynamics and Proton Transfer within the Adenine-Thymine Nucleobase Pair. *J. Phys. Chem. Lett.* **2023**, *14*, 3698–3705.

(14) Rangadurai, A.; Szymanski, E. S.; Kimsey, I.; Shi, H.; Al-Hashimi, H. M. Probing Conformational Transitions towards Mutagenic Watson–Crick-like G·T Mismatches Using off-Resonance Sugar Carbon R1ρ Relaxation Dispersion. *J Biomol NMR* **2020**, *74*, 457–471.

(15) Botti, G.; Ceotto, M.; Conte, R. On-the-Fly Adiabatically Switched Semiclassical Initial Value Representation Molecular Dynamics for Vibrational Spectroscopy of Biomolecules. *J. Chem. Phys.* **2021**, *155*, 234102.

(16) Botti, G.; Aieta, C.; Conte, R. The Complex Vibrational Spectrum of Proline Explained through the Adiabatically Switched Semiclassical Initial Value Representation. *J. Chem. Phys.* **2022**, *156*, 164303.

(17) Fischer, T. L.; Wagner, T.; Gottschalk, H. C.; Nejad, A.; Suhm, M. A. A Rather Universal Vibrational Resonance in 1:1 Hydrates of Carbonyl Compounds. *J. Phys. Chem. Lett.* **2021**, *12*, 138–144.

(18) Kabeláč, M.; Ryjáček, F.; Hobza, P. Already Two Water Molecules Change Planar H-bonded Structures of the Adenine···thymine Base Pair to the Stacked Ones: A Molecular Dynamics Simulations Study. *Phys. Chem. Chem. Phys.* **2000**, *2*, 4906–4909.

(19) Seki, T.; Chiang, K.-Y.; Yu, C.-C.; Yu, X.; Okuno, M.; Hunger, J.; Nagata, Y.; Bonn, M. The Bending Mode of Water: A Powerful Probe for Hydrogen Bond Structure of Aqueous Systems. *J. Phys. Chem. Lett.* **2020**, *11*, 8459–8469.

(20) Fecko, C. J.; Eaves, J. D.; Loparo, J. J.; Tokmakoff, A.; Geissler, P. L. Ultrafast Hydrogen-Bond Dynamics in the Infrared Spectroscopy of Water. *Science* **2003**, *301*, 1698–1702.

(21) Dean, J. L. S.; Fournier, J. A. Vibrational Dynamics of the Intramolecular H-Bond in Acetylacetone Investigated with Transient and 2D IR Spectroscopy. *J. Phys. Chem. B* **2022**, *11*, 4.

(22) Biswal, H. S.; Gloaguen, E.; Loquais, Y.; Tardivel, B.; Mons, M. Strength of NH$\cdots$S Hydrogen Bonds in Methionine Residues Revealed by Gas-Phase IR/UV Spectroscopy. *J. Phys. Chem. Lett.* **2012**, *3*, 755–759.

(23) Morawietz, T.; Urbina, A. S.; Wise, P. K.; Wu, X.; Lu, W.; Ben-Amotz, D.; Markland, T. E. Hiding in the Crowd: Spectral Signatures of Overcoordinated Hydrogen-Bond Environments. *J. Phys. Chem. Lett.* **2019**, *10*, 6067–6073.

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

(25) Gabas, F.; Di Liberto, G.; Conte, R.; Ceotto, M. Protonated Glycine Supramolecular Systems: The Need for Quantum Dynamics. *Chem. Sci.* **2018**, *9*, 7894–7901.

(26) Bertaina, G.; Di Liberto, G.; Ceotto, M. Reduced Rovibrational Coupling Cartesian Dynamics for Semiclassical Calculations: Application to the Spectrum of the Zundel Cation. *J. Chem. Phys.* **2019**, *151*, 114307.

(27) Fallacara, E.; Depondt, P.; Huppert, S.; Ceotto, M.; Finocchi, F. Thermal and Nuclear Quantum Effects at the Antiferroelectric to Paraelectric Phase Transition in KOH and KOD Crystals. *J. Phys. Chem. C* **2021**, *125*, 22328–22334.

(28) Fischer, T. L.; Bödecker, M.; Schweer, S. M.; Dupont, J.; Lepère, V.; Zehnacker-Rentien, A.; Suhm, M. A.; Schröder, B.; Henkes, T.; Andrada, D. M. et al. The First HyDRA Challenge for Computational Vibrational Spectroscopy. *Phys. Chem. Chem. Phys.* **2023**, *25*, 22089–22102.

(29) Abo-Riziq, A.; Grace, L.; Nir, E.; Kabelac, M.; Hobza, P.; de Vries, M. S. Photochemical Selectivity in Guanine–Cytosine Base-Pair Structures. *Proc. Natl. Acad. Sci.* **2005**, *102*, 20–23.

(30) Kabelac, M.; Hobza, P. At Nonzero Temperatures, Stacked Structures of Methylated Nucleic Acid Base Pairs and Microhydrated Nonmethylated Nucleic Acid Base Pairs Are Favored over Planar Hydrogen-Bonded Structures: A Molecular Dynamics Simulations Study. *Chemistry A European J* **2001**, *7*, 2067–2074.

(31) Nir, E.; Janzen, C.; Imhof, P.; Kleinermanns, K.; de Vries, M. S. Guanine Tautomerism Revealed by UV–UV and IR–UV Hole Burning Spectroscopy. *J. Chem. Phys.* **2001**, *115*, 4604–4611.

(32) Nir, E.; Janzen, C.; Imhof, P.; Kleinermanns, K.; de Vries, M. S. Pairing of the Nucleobases Guanine and Cytosine in the Gas Phase Studied by IR–UV Double-Resonance Spectroscopy and Ab Initio Calculations. *Phys. Chem. Chem. Phys.* **2002**, *4*, 732–739.

(33) Nir, E.; Plützer, C.; Kleinermanns, K.; de Vries, M. Properties of Isolated DNA Bases, Base Pairs and Nucleosides Examined by Laser Spectroscopy. *Eur. Phys. J. D* **2002**, *20*, 317–329.

(34) Bakker, J. M.; Compagnon, I.; Meijer, G.; von Helden, G.; Kabeláč, M.; Hobza, P.; de Vries, M. S. The Mid-IR Absorption Spectrum of Gas-Phase Clusters of the Nucleobases Guanine and Cytosine. *Phys. Chem. Chem. Phys.* **2004**, *6*, 2810–2815.

(35) Plützer, C.; Hünig, I.; Kleinermanns, K.; Nir, E.; de Vries, M. S. Pairing of Isolated Nucleobases: Double Resonance Laser Spectroscopy of Adenine-Thymine. *ChemPhysChem* **2003**, *4*, 838–842.

(36) Plützer, C.; Hünig, I.; Kleinermanns, K. Pairing of the Nucleobase Adenine Studied by IR-UV Double-Resonance Spectroscopy and Ab Initio Calculations. *Phys. Chem. Chem. Phys.* **2003**, *5*, 1158–1163.

(37) Brauer, B.; Gerber, R. B.; Kabeláč, M.; Hobza, P.; Bakker, J. M.; Abo Riziq, A. G.; de Vries, M. S. Vibrational Spectroscopy of the G$\cdots$C Base Pair: Experiment, Harmonic and Anharmonic Calculations, and the Nature of the Anharmonic Couplings. *J. Phys. Chem. A* **2005**, *109*, 6974–6984.

(38) Sobolewski, A. L.; Domcke, W. Ab Initio Studies on the Photophysics of the Guanine–Cytosine Base Pair. *Phys. Chem. Chem. Phys.* **2004**, *6*, 2763–2771.

(39) Jiang, Y.; Wang, L. Development of Vibrational Frequency Maps for Nucleobases. *J. Phys. Chem. B* **2019**, *123*, 5791–5804.

(40) Fornaro, T.; Biczysko, M.; Monti, S.; Barone, V. Dispersion Corrected DFT Approaches for Anharmonic Vibrational Frequency Calculations: Nucleobases and Their Dimers. *Phys. Chem. Chem. Phys.* **2014**, *16*, 10112–10128.

(41) 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.

(42) 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.

(43) Di Liberto, G.; Conte, R.; Ceotto, M. "Divide-and-conquer" Semiclassical Molecular Dynamics: An Application to Water Clusters. *J. Chem. Phys.* **2018**, *148*, 104302.

(44) 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. *J. Chem. Phys.* **2020**, *153*, 204104.

(45) Gabas, F.; Di Liberto, G.; Ceotto, M. Vibrational Investigation of Nucleobases by Means of Divide and Conquer Semiclassical Dynamics. *J. Chem. Phys.* **2019**, *150*, 224107.

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

(47) Miller, W. H. Semiclassical Theory of Atom–Diatom Collisions: Path Integrals and the Classical S Matrix. *J. Chem. Phys.* **1970**, *53*, 1949–1959.

(48) Elran, Y.; Kay, K. G. Improving the Efficiency of the Herman–Kluk Propagator by Time Integration. *J. Chem. Phys.* **1999**, *110*, 3653–3659.

(49) Elran, Y.; Kay, K. G. Time-Integrated Form of the Semiclassical Initial Value Method. *J. Chem. Phys.* **1999**, *110*, 8912–8918.

(50) Kaledin, A. L.; Miller, W. H. Time Averaging the Semiclassical Initial Value Representation for the Calculation of Vibrational Energy Levels. II. Application to H2CO, NH3, CH4, CH2D2. *J. Chem. Phys.* **2003**, *119*, 3078–3084.

(51) Herman, M. F.; Kluk, E. A Semiclasical Justification for the Use of Non-Spreading Wavepackets in Dynamics Calculations. *Chem. Phys.* **1984**, *91*, 27–34.

(52) 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.

(53) Ceotto, M.; Atahan, S.; Shim, S.; Tantardini, G. F.; Aspuru-Guzik, A. First-Principles Semiclassical Initial Value Representation Molecular Dynamics. *Phys. Chem. Chem. Phys.* **2009**, *11*, 3861.

(54) Conte, R.; Aspuru-Guzik, A.; Ceotto, M. Reproducing Deep Tunneling Splittings, Resonances, and Quantum Frequencies in Vibrational Spectra From a Handful of Direct Ab Initio Semiclassical Trajectories. *J. Phys. Chem. Lett.* **2013**, *4*, 3407–3412.

(55) Gabas, F.; Conte, R.; Ceotto, M. On-the-Fly Ab Initio Semiclassical Calculation of Glycine Vibrational Spectrum. *J. Chem. Theory Comput.* **2017**, *13*, 2378–2388.

(56) De Leon, N.; Heller, E. J. Semiclassical Quantization and Extraction of Eigenfunctions Using Arbitrary Trajectories. *J. Chem. Phys.* **1983**, *78*, 4005–4017.

(57) 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.

(58) Rognoni, A.; Conte, R.; Ceotto, M. Caldeira–Leggett Model vs *Ab Initio* Potential: A Vibrational Spectroscopy Test of Water Solvation. *J. Chem. Phys.* **2021**, *154*, 094106.

(59) Rognoni, A.; Conte, R.; Ceotto, M. How Many Water Molecules Are Needed to Solvate One? *Chem. Sci.* **2021**, *12*, 2060–2064.

(60) Weigend, F.; Ahlrichs, R. Balanced Basis Sets of Split Valence, Triple Zeta Valence and Quadruple Zeta Valence Quality for H to Rn: Design and Assessment of Accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297.

(61) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate *Ab Initio* Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.

(62) Angiolari, F.; Huppert, S.; Pietrucci, F.; Spezia, R. Environmental and Nuclear Quantum Effects on Double Proton Transfer in the Guanine–Cytosine Base Pair. *J. Phys. Chem. Lett.* **2023**, *14*, 5102–5108.