Supporting Information: Unraveling Water Solvation Effects with Quantum Mechanics/Molecular Mechanics Semiclassical Vibrational Spectroscopy: The case of Thymidine

Davide Moscato,[†] Giacomo Mandelli,[†] Mattia Bondanza,[‡] Filippo Lipparini,[*,‡] Riccardo Conte,[†] Benedetta Mennucci,[‡] and Michele Ceotto[*,†]

[†]*Dipartimento di Chimica, Università degli Studi di Milano, Via Golgi, 19 - 20133 Milano - Italy*

[‡]*Dipartimento di Chimica e Chimica Industriale, Università di Pisa, Via Giuseppe Moruzzi, 13 - 56124 Pisa - Italy*

E-mail: filippo.lipparini@unipi.it; michele.ceotto@unimi.it

The Semiclassical Initial Value Representation

The inclusion of Nuclear Quantum Effects (NQEs) in the evaluation of spectroscopical features is not a trivial task, but in some cases can be fundamental for the sake of a complete description of the system. Within a quantum dynamical approach to vibrational spectroscopy, observing anharmonic overtones, combination bands, Fermi's resonances, anharmonic Zero Point Energy (ZPE), and tunneling splittings is possible. In this framework, the power spectra of a system composed of N degrees of freedom with Hamiltonian $\hat{H}$ is defined as the Fourier Transform (FT) of the surviving amplitude of an arbitrary reference state $|\Psi\rangle$

$$I(E) = \frac{1}{2\pi\hbar} \int_{-\infty}^{+\infty} \langle \Psi \Big| e^{-\frac{i}{\hbar}\hat{H}t} \Big| \Psi \rangle \, e^{\frac{i}{\hbar}Et} \, dt. \tag{S1}$$

Equation S1 can be evaluated exactly only for simple model systems. Other "numerically exact" methods can make highly accurate estimates for small molecular systems.[1–3] Another way to evaluate eq. S1, is to find an approximate form of the time evolution operator ($e^{-\frac{i}{\hbar}\hat{H}t}$). In Feynman's Path Integral (PI) representation of the propagator, the total amplitude for a system evolving from a point in space $\mathbf{q}(0)$ to $\mathbf{q}(t)$, is given by considering all of the possible infinite paths that connect the two points in space.[4]

$$\langle \mathbf{q}(t) \big| e^{-\frac{i}{\hbar}\hat{H}t} \big| \mathbf{q}(0) \rangle = \int_{\mathbf{q}(0)}^{\mathbf{q}(t)} \mathcal{D}[\mathbf{q}] \; e^{\frac{i}{\hbar}S(\mathbf{q}(t))} \; , \tag{S2}$$

where $\mathcal{D}[\mathbf{q}]$ includes the PI prefactor which accounts for the weight of every path and $S(\mathbf{q}(t))$ is the action along the considered path. Most of the paths included in the exact PI representation of the propagator are non-classical. It is anyway possible to apply a stationary phase approximation to the PI propagator, so that only paths which make the action stationary (classical paths) are considered. Quantum mechanical effects retained as second-order deviations of the action and also by the sum over classical paths. This kind of approximation of the PI is also known as the semiclassical (SC) approximation[5,6]

$$\langle \mathbf{q}(t) \Big| e^{-\frac{i}{\hbar}\hat{H}t} \Big| \mathbf{q}(0) \rangle \approx \sum_{roots} \sqrt{\frac{1}{(2\pi\hbar)^N} \left| \frac{\partial \mathbf{q}(t)}{\partial \mathbf{p}(0)} \right|^{-1}} \, e^{\frac{i}{\hbar} S(\mathbf{q}(t), \mathbf{q}(0))} \ , \tag{S3}$$

where $\left| \frac{\partial \mathbf{q}(t)}{\partial \mathbf{p}(0)} \right|$ is the determinant of the matrix obtained from the second-order deviations of the action. The sum of equation S3 runs over the "roots" which are the paths that make the action functional stationary, $S(\mathbf{q}(t), \mathbf{q}(0))$ is the classical action between the points in space $\mathbf{q}(0)$ and $\mathbf{q}(t)$

$$S(\mathbf{q}(t), \mathbf{q}(0)) = \int_0^t \Big[\mathbf{p}(\tau)\dot{\mathbf{q}}(\tau) - H(\mathbf{p}(\tau), \mathbf{q}(\tau)) \Big] \, d\tau. \tag{S4}$$

In this way, quantum effects are reproduced by considering all the classical trajectories that connect two points in space in a given time. In a practical way, this approximation of the propagator is problematic for two reasons. The most evident is related to the fact that in multidimensional cases the double boundary condition becomes more and more challenging to satisfy. The other issue arises in the so-called caustic points where $\left| \frac{\partial \mathbf{q}(t)}{\partial \mathbf{p}(0)} \right| = 0$ and the prefactor matrix becomes non-invertible. A fundamental step in the development of the semiclassical propagator was carried out by Miller, which led to the Initial Value Representation (IVR).[7] In this way the double boundary problem was eased up and the problem of caustic points was solved. Further developments of the semiclassical propagator were perpetrated by Heller, Herman, Kay, and Kluk[8–12] (HHKK) in different instances to solve the issue of the double boundary conditions and rewrote the propagator in terms of coherent states in phase-space representation. In this way, initial conditions are sampled with an importance Monte Carlo (MC) sampling based on the Husimi distribution

$$e^{-\frac{i}{\hbar}\hat{H}t}_{HHKK} = \frac{1}{(2\pi\hbar)^N} \iint d\mathbf{p}(0) d\mathbf{q}(0) \; C\left(\mathbf{p}(0), \mathbf{q}(0), t\right) \; e^{\frac{i}{\hbar} S(\mathbf{p}(0), \mathbf{q}(0), t)} \left| \mathbf{p}(t), \mathbf{q}(t) \right\rangle \left\langle \mathbf{p}(0), \mathbf{q}(0) \right| \ , \tag{S5}$$

where $C\left(\mathbf{p}(0), \mathbf{q}(0), t\right)$ is the HHKK prefactor defined as

$$C(\mathbf{p}(0),\mathbf{q}(0),t)=\sqrt{\left|\frac{1}{2}\left(\frac{\partial\mathbf{q}(t)}{\partial\mathbf{q}(0)}+\Gamma^{-1}\frac{\partial\mathbf{p}(t)}{\partial\mathbf{p}(0)}\Gamma-i\hbar\frac{\partial\mathbf{q}(t)}{\partial\mathbf{p}(0)}\Gamma+\frac{i\Gamma^{-1}}{\hbar}\frac{\partial\mathbf{p}(t)}{\partial\mathbf{q}(0)}\right)\right|}, \tag{S6}$$

which contains the block of the monodromy matrix of the system. $|\mathbf{p}(t),\mathbf{q}(t)\rangle$ and $\langle\mathbf{p}(0),\mathbf{q}(0)|$ are coherent states for which

$$\langle\mathbf{x}\big|\mathbf{p}(t),\mathbf{q}(t)\rangle=\left(\frac{\det(\Gamma)}{\pi^N}\right)^{1/4}e^{-\frac{1}{2}\left(\mathbf{x}-\mathbf{q}(t)\right)\Gamma\left(\mathbf{x}-\mathbf{q}(t)\right)+\frac{i}{\hbar}\mathbf{p}(t)\left(\mathbf{x}-\mathbf{q}(t)\right)}\ , \tag{S7}$$

where Γ is the width of the Gaussian, which in our case is a diagonal matrix with non-zero elements equal to the harmonic frequencies of the system. Finally $S(\mathbf{p}(0),\mathbf{q}(0),t)$ is the classical action evaluated along the trajectory originated by the phase-space point $(\mathbf{p}(0),\mathbf{q}(0))$.

The practical application of the HHKK propagator, see equation (S5), still presents issues when dealing with molecular systems. A problem comes from the fact that the number of MC samples required to achieve convergence scales exponentially with the number of degrees of freedom due to the strongly oscillatory behavior of the integrand. The second issue instead arises from the pre-exponential factor, see equation (S6), introduced in the HHKK representation. In fact, some of the trajectories exhibit a "chaotic" behavior where this term displays exponential growth. For this reason, trajectories displaying this chaotic behavior needs to be rejected making the MC convergence even harder to achieve. Starting from the HHKK propagator, for equation (S1) Kaledin and Miller proposed a time averaging filter[13–16] (TA SCIVR) in order to ease the MC convergence issues

$$\begin{aligned} I(E)=\frac{1}{(2\pi\hbar)^N}\iint d\mathbf{p}(0)d\mathbf{q}(0)\\ \frac{1}{2\pi\hbar T}\left|\int_0^T dt\ \langle\Psi\big|\mathbf{p}(t),\mathbf{q}(t)\rangle\ e^{\frac{i}{\hbar}\left[S(\mathbf{p}(0),\mathbf{q}(0),t)+Et+\phi(\mathbf{p}(0),\mathbf{q}(0),t)\right]}\right|^2, \end{aligned} \tag{S8}$$

where $\langle\Psi\big|\mathbf{p}(t),\mathbf{q}(t)\rangle$ is the overlap between the reference state which is usually chosen to be a coherent state, and the Gaussian coherent state centered at the phase space point $(\mathbf{p}(t),\mathbf{q}(t))$. $\phi(\mathbf{p}(0),\mathbf{q}(0),t)$ is the phase of the HHKK pre-exponential factor. In this way the phase space integrand is a positive defined real function and this speeds up the MC convergence. The evaluation of the monodromy matrix elements still represents a huge computational bottleneck to the whole family of semiclassical methods, since it requires the computation of the Hessian matrices of the potential along the trajectories.[17,18] Within this framework, the computation of vibrational spectra of small and medium-sized molecular systems becomes possible if a fitted PES is available. Indeed, the number of trajectories and therefore of Hessian matrices required to achieve convergence is still prohibitive for *ab initio* on-the-fly calculations. To solve this problem the multiple coherent (MC) SCIVR[19,20] method was developed. This method is based on a pivotal work published by Heller and De Leon[21] in which it was demonstrated that with a single trajectory run at the "correct" energy it is possible to obtain accurate semiclassical eigenenergies and eigenfunctions. In MC SCIVR the assumption made is that a reliable frequency calculation for molecular systems can be made by means of a single trajectory run near the actual but unknown energy. The easiest and most intuitive way to run trajectories near the true quantum mechanical values is to start from quantized harmonic conditions

$$\begin{cases} p_i(0) &= \sqrt{\hbar\omega_i(2n_i+1)} \\ q_i(0) &= q_i^{(eq)} \end{cases}, \tag{S9}$$

where the initial momentum in mass-scaled coordinates for the normal mode i is proportional to the square root of the corresponding harmonic frequency ω_i and the vibrational quantum number n_i. The initial position is chosen to be equal to the equilibrium geometry. Another feature that makes the MC-SCIVR method convenient for spectroscopic calculations of large molecular systems is the choice of the reference state, which in this method is a basis set composed of Gaussian coherent states

$$|\Psi^{(K)}\rangle = \prod_{J=1}^{N} \left(\varepsilon_{1,J} \, |p_J^{(K)}(0), q_J^{(K)}(0)\rangle + \varepsilon_{2,J} \, |-p_J^{(K)}(0), q_J^{(K)}(0)\rangle \right) \ . \tag{S10}$$

By choosing the value of epsilon as $+1, -1$ or 0 for different elements of the base it is possible to enhance the signal of the desired normal modes in the power spectrum. In MC SCIVR the integration over the phase space is substituted with a summation over all the trajectories

$$I(E) = \frac{1}{(2\pi\hbar)^N} \sum_{K}^{N_{traj}} \frac{1}{2\pi\hbar T} \left| \int_0^T dt \ e^{\frac{i}{\hbar}\left[S(\mathbf{p}^{(K)}(0),\mathbf{q}^{(K)}(0),t)+Et+\phi(\mathbf{p}^{(K)}(0),\mathbf{q}^{(K)}(0),t)\right]} \, \langle \Psi^{(K)} \big| \mathbf{p}^{(K)}(t), \mathbf{q}^{(K)}(t) \rangle \right|^2 \ . \tag{S11}$$

Even though this step of development permits on-the-fly computation of quantum mechanical power spectra, it is still problematic to obtain well resolved and intense signal from systems with a large number of degrees of freedom. The problem derives from the fact that the overlap between the coherent state and the reference state is actually a product of overlap integrals between coherent states of all degrees of freedom featured in the system

$$\begin{aligned} \langle \Psi \big| \mathbf{p}(t), \mathbf{q}(t) \rangle = & \\ & \left(\varepsilon_{1,1} \, \langle p_1(0), q_1(0) \big| p_1(t), q_1(t) \rangle + \varepsilon_{2,1} \, \langle -p_1(0), q_1(0) \big| p_1(t), q_1(t) \rangle \right) \times \\ & \left(\varepsilon_{1,2} \, \langle p_2(0), q_2(0) \big| p_2(t), q_2(t) \rangle + \varepsilon_{2,2} \, \langle -p_2(0), q_2(0) \big| p_2(t), q_2(t) \rangle \right) \times \\ & \vdots \\ & \times \left(\varepsilon_{1,N} \, \langle p_N(0), q_N(0) \big| p_N(t), q_N(t) \rangle + \varepsilon_{2,N} \, \langle -p_N(0), q_N(0) \big| p_N(t), q_N(t) \rangle \right) \ . \end{aligned} \tag{S12}$$

By observing equation (S12), it becomes evident that as the number of degrees of freedom increases, the likelihood of one of the elements in the product approaching zero also increases.

This will cause the whole overlap integral to have a value near to zero, causing the spectrum to be noisy and with poor resolution. A divide and conquer (DC SCIVR) approach to the semiclassical method was proposed to overcome this problem.[22,23] The working formula in the multiple coherent case is the following

$$\tilde{I}(E) = \frac{1}{(2\pi\hbar)^M} \sum_{K}^{N_{traj}} \frac{1}{2\pi\hbar T} \left| \int_0^T dt\, e^{\frac{i}{\hbar}\left[\tilde{S}\left(\tilde{\mathbf{p}}^{(K)}(0),\tilde{\mathbf{q}}^{(K)}(0),t\right)+Et+\tilde{\phi}\left(\tilde{\mathbf{p}}^{(K)}(0),\tilde{\mathbf{q}}^{(K)}(0),t\right)\right]} \left\langle \tilde{\Psi}^{(K)} \middle| \tilde{\mathbf{p}}^{(K)}(t), \tilde{\mathbf{q}}^{(K)}(t) \right\rangle \right|^2 . \tag{S13}$$

The main difference between this formula and the original one of equation (S11) is that the dynamical quantities in equation (S13) are projected onto M-dimensional subspaces. In practice with this approach a full dimensional classical trajectory is evolved on the true potential, while the evaluation of the monodromy matrix elements and the overlap between coherent states is made only on the normal modes of the selected subspaces. The projection of the action is convoluted, since in complex systems the potential is non-separable. A working approximation to solve this problem comes from considering a trivial case where the potential depends directly on the M degrees of freedom of the selected subspace. The remaining $N-M$ degrees of freedom are treated as parameters, and, in addition, to account for the non-separability of the potential, a time-dependent external field is summed to the first term

$$\tilde{V}(\mathbf{q}_M(t)) = V(\mathbf{q}_M(t); \mathbf{q}^{eq}_{N-M}) + \lambda(t) \ , \tag{S14}$$

where $\tilde{V}(\mathbf{q}_M(t))$ is the projected potential and $\lambda(t)$ is the external field.

$$\lambda(t) = V(\mathbf{q}_M(t); \mathbf{q}_{N-M}(t)) - V(\mathbf{q}_M(t); \mathbf{q}^{eq}_{N-M}) - V(\mathbf{q}^{eq}_M; \mathbf{q}_{N-M}(t)) \ . \tag{S15}$$

By representing the projected potential as in equations S14 and S15, the approximation is exact in the separable case. The dimensions and content of the subspaces are chosen by

evaluating how much normal modes are coupled with each other.[24]

Computational details

The thymidine geometry was first optimized using Gaussian 16[25] at the B3LYP/6-31g*[26–31] level of theory with Grimme's dispersion corrections (GD3).[32] A sphere with 15 Å radius filled with water molecules was generated using the Packmol tool.[33] The number of water molecules was chosen to reproduce the density of liquid water. With the TINKER[34] software suite repulsive walls were put on the edge of the solvation sphere, and the optimization of the system was done. Subsequently, the thymidine optimized geometry was inserted in the water sphere, and an optimization with the AMOEBABIO18[35] force field (FF) of the water molecules was carried on. The QM/MM calculations were made using an in-house modified version of the Gaussian software which is interfaced with TINKER. In this case, a further full optimization is carried on. The numerical equilibrium Hessian matrix is then computed and diagonalized to identify the vibrational normal modes of interest. Initial conditions for the dynamics were chosen according to equation (S9), with n_i set equal to zero for all the normal modes. A 3000-step NVE trajectory with 0.2 fs timestep was propagated, and Hessian matrices along the trajectory were calculated every 100 steps. With the calculated matrices along the trajectory, an average matrix was evaluated in order to divide the system into vibrational subspaces. The elements of the monodromy matrix along the trajectory are then calculated by using a finite difference scheme in which only the required elements are considered.[36] The semiclassical vibrational spectra were then computed according to equation (S13). For the PCM calculations, the optimization of the thymidine molecule was done at the same level of theory used in the simulation with the explicit solvent, with the addition of the PCM solvent implemented in Gaussian 16. The analytical Hessian matrix at the equilibrium configuration was used to compute and identify the vibrational normal modes. The NVE MD simulation was then carried out with the same software and the

same number of steps and time-step as before. The initial conditions are set in the same way as the QM/MM run. In this case, the second derivative matrices along the trajectory were evaluated at each step.

Table S1: Summary the calculated frequencies for the different levels of theory in cm^{-1}

	Harmonic			
	Str. C5C6	Str. C4O	MAE	Δ
B3LYP@6-31g*/AMOEBA	1720	1738	20	18
B3LYP@6-31g*/AMOEBA (no pol)	1711	1767	29	87
B3LYP@6-31g*/TIP3P	1719	1806	52	87
B3LYP@6-31g*/PCM	1709	1743	11	34
AMOEBABIO18	1619	1664	69	45
	Quasiclassical Trajectory			
	Str. C5C6	Str. C4O	MAE	Δ
B3LYP@6-31g*/AMOEBA	1684	1695	20	11
B3LYP@6-31g*/AMOEBA (no pol)	1663	1748	42	85
B3LYP@6-31g*/TIP3P	1691	1780	44	89
B3LYP@6-31g*/PCM	1699	1722	11	23
AMOEBABIO18	1612	1618	95	5
	MC DC SCIVR			
	Str. C5C6	Str. C4O	MAE	Δ
B3LYP@6-31g*/AMOEBA	1689	1701	15	12
B3LYP@6-31g*/AMOEBA (no pol)	1660	1748	44	88
B3LYP@6-31g*/TIP3P	1687	1791	52	104
B3LYP@6-31g*/PCM	1660	1748	5	10
AMOEBABIO18	1621	1634	82	13
	Experiment			
	Str. C5C6		Str. C4O	
	1710		1710	

Analysis of the trajectories

The different spectroscopic signals obtained by fine-tuning the solvent model can be rationalized by means of trajectory analysis. To analyze how the solvent arranges around specific areas of the molecule it is possible to calculate the Radial Distribution Function (RDF). In this work, we are interested in observing the behavior of two normal modes, the $C4O$ and $C5C6$ stretches. One of them is strongly localized in the $C4$ carboxylic group and the other

between the $C5$ and $C6$ atoms. Therefore, the average RDF[37] was calculated both between the $C5 - C6$ and $C4 - O$ atoms and all of the water molecules.

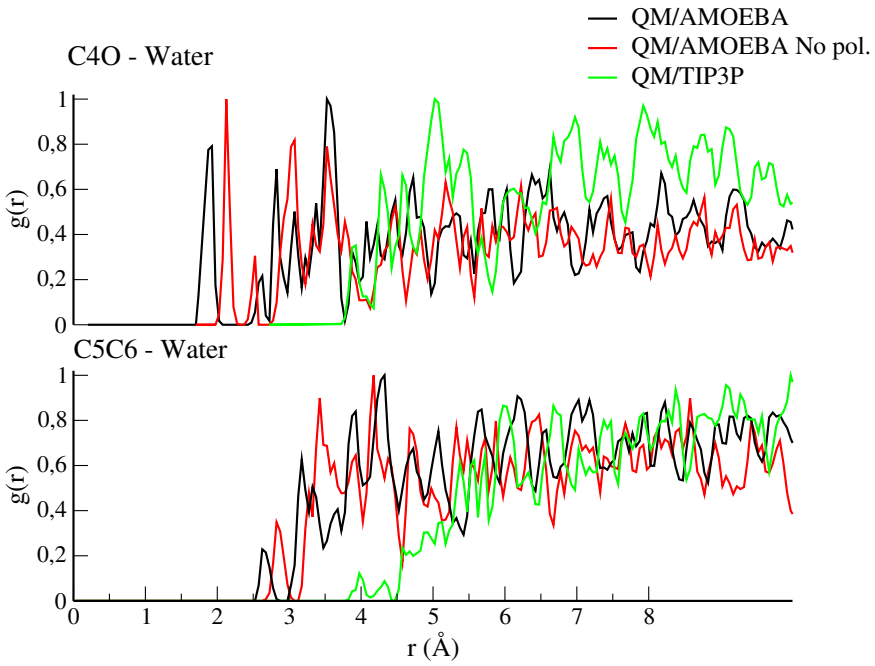

Figure S1: Radial Distribution Function (RDF) calculated between $C4 - O$ and water molecules (upper panel) and between $C5 - C6$ and water molecules (lower panel). The calculations have been carried out for all of the tested QM/MM approaches.

By observing the calculated RDFs in Figure S1, it is possible to observe that by removing the polarization from the MM part, the water molecules tend to stay more distant from the thymidine molecule. By impoverishing the electrostatic description of the solvent, like in the QM/TIP3P case, it is clear that the solvent molecules tend to be even more distanced with respect to a multipolar description like in the QM/AMOEBA case with no polarization included. This could be one of the reasons why the non polarizable QM/MM models reproduce a spectroscopic behavior similar to the isolated thymidine molecule.

The PCM results are in good agreement compared to both the experimental data and the polarizable QM/AMOEBA approach. This leads us to think that no significant directional interactions between the solvent and the central molecule occur. To verify this, a search for hydrogen bond interactions between the solvent molecules and the thymidine molecule was

made along the QM/AMOEBA trajectory using the MDAnalysis library.[38,39] The search has been made for two cases when the thymidine atoms serve as acceptors (O and N atoms involved) and water acts as a donor (H atoms involved), and vice versa. A first search for interactions was made by setting a distance cutoff of 2.5 Å and an angle cutoff of 170°, which means that all of the non-bonded interactions between the selected group of atoms with a distance greater than the cutoff are automatically excluded. The angle cutoff excludes all the non-bonded interactions where the donor-hydrogen-acceptor angle value is below the cutoff. This first search led to zero interactions found and for this reason, less severe parameters were tested. It was observed that by elongating the distance cutoff and keeping the angle cutoff the same still no interactions were found. Finally by setting the distance cutoff to 3.0 Å and the angle cutoff to 120° some interactions were observed.

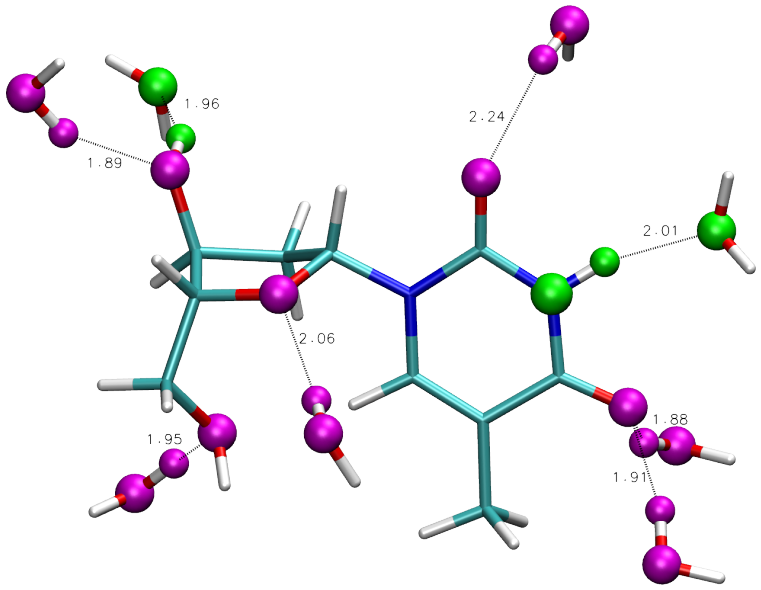

Figure S2: Picture of the thymidine molecule in the minimum geometry where hydrogen bonds are highlighted as purple for acceptors and green for donors

This set of search parameters actually corroborated the hypothesis that the interactions are not strongly directional. Therefore, we can affirm that the PCM results are not due to compensations of error and can be justified. Moreover, thanks to this analysis we are able to conclude that the degeneration of the two Raman signals observed in the experiment is mainly due to dispersive interactions of the molecule with the solvent.

References

(1) Light, J. C.; Bacic, Z. Adiabatic approximation and nonadiabatic corrections in the discrete variable representation: Highly excited vibrational states of triatomic molecules. *The Journal of Chemical Physics* **1987**, *87*, 4008–4019.

(2) Carter, S.; Pinnavaia, N.; Handy, N. C. The vibrations of formaldehyde. *Chemical Physics Letters* **1995**, *240*, 400–408.

(3) Wang, X.-G.; Carrington, J., Tucker Computing excited OH stretch states of water dimer in 12D using contracted intermolecular and intramolecular basis functions. *The Journal of Chemical Physics* **2023**, *158*, 084107.

(4) Feynman, R. P. Space-Time Approach to Non-Relativistic Quantum Mechanics. *Rev. Mod. Phys.* **1948**, *20*, 367–387.

(5) Vleck, J. H. V. The Correspondence Principle in the Statistical Interpretation of Quantum Mechanics. *Proceedings of the National Academy of Sciences* **1928**, *14*, 178–188.

(6) Conte, R.; Ceotto, M. *Quantum Chemistry and Dynamics of Excited States*; John Wiley and Sons, Ltd, 2020; Chapter 19, pp 595–628.

(7) Miller, W. H. Classical S Matrix: Numerical Application to Inelastic Collisions. *The Journal of Chemical Physics* **2003**, *53*, 3578–3587.

(8) Heller, E. J. Frozen Gaussians: A very simple semiclassical approximation. *The Journal of Chemical Physics* **1981**, *75*, 2923–2931.

(9) Herman, M. F.; Kluk, E. A semiclasical justification for the use of non-spreading wavepackets in dynamics calculations. *Chemical Physics* **1984**, *91*, 27–34.

(10) Kay, K. G. Integral expressions for the semiclassical time dependent propagator. *The Journal of Chemical Physics* **1994**, *100*, 4377–4392.

(11) Kay, K. G. Numerical study of semiclassical initial value methods for dynamics. *The Journal of Chemical Physics* **1994**, *100*, 4432–4445.

(12) Kay, K. G. Semiclassical propagation for multidimensional systems by an initial value method. *The Journal of Chemical Physics* **1994**, *101*, 2250–2260.

(13) Elran, Y.; Kay, K. G. Improving the efficiency of the Herman-Kluk propagator by time integration. *The Journal of Chemical Physics* **1999**, *110*, 3653–3659.

(14) Elran, Y.; Kay, K. G. Time-integrated form of the semiclassical initial value method. *The Journal of Chemical Physics* **1999**, *110*, 8912–8918.

(15) Kaledin, A. L.; Miller, W. H. Time averaging the semiclassical initial value representation for the calculation of vibrational energy levels. *The Journal of Chemical Physics* **2003**, *118*, 7174–7182.

(16) 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. *The Journal of Chemical Physics* **2003**, *119*, 3078–3084.

(17) Conte, R.; Gabas, F.; Botti, G.; Zhuang, Y.; Ceotto, M. Semiclassical vibrational spectroscopy with Hessian databases. *The Journal of Chemical Physics* **2019**, *150*, 244118.

(18) Gandolfi, M.; Ceotto, M. Unsupervised Machine Learning Neural Gas Algorithm for Accurate Evaluations of the Hessian Matrix in Molecular Dynamics. *Journal of Chemical Theory and Computation* **2021**, *17*, 6733–6746.

(19) Ceotto, M.; Atahan, S.; Tantardini, G. F.; Aspuru-Guzik, A. Multiple coherent states for first-principles semiclassical initial value representation molecular dynamics. *The Journal of Chemical Physics* **2009**, *130*, 234113.

(20) 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*, 3861–3867.

(21) De Leon, N.; Heller, E. J. Semiclassical quantization and extraction of eigenfunctions using arbitrary trajectories. *The Journal of Chemical Physics* **1983**, *78*, 4005–4017.

(22) Ceotto, M.; Di Liberto, G.; Conte, R. Semiclassical "Divide-and-Conquer" method for spectroscopic calculations of high dimensional molecular systems. *Physical Review Letters* **2017**, *119*, 010401.

(23) Di Liberto, G.; Conte, R.; Ceotto, M. "Divide and conquer" semiclassical molecular dynamics: A practical method for spectroscopic calculations of high dimensional molecular systems. *The Journal of Chemical Physics* **2018**, *148*, 014307.

(24) 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.

(25) Frisch, M. e.; Trucks, G.; Schlegel, H. B.; Scuseria, G.; Robb, M.; Cheeseman, J.; Scalmani, G.; Barone, V.; Petersson, G.; Nakatsuji, H.; others Gaussian 16. 2016.

(26) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Physical review A* **1988**, *38*, 3098.

(27) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical review B* **1988**, *37*, 785.

(28) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *The Journal of physical chemistry* **1994**, *98*, 11623–11627.

(29) Ditchfield, R.; Hehre, W. J.; Pople, J. A. Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1971**, *54*, 724–728.

(30) Hehre, W. J.; Ditchfield, R.; Pople, J. A. Self-Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian-Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1972**, *56*, 2257–2261.

(31) Hariharan, P. C.; Pople, J. A. The influence of polarization functions on molecular orbital hydrogenation energies. *Theor. Chim. Acta* **1973**, *28*, 213–222.

(32) Ehrlich, S.; Moellmann, J.; Reckien, W.; Bredow, T.; Grimme, S. System-dependent dispersion coefficients for the DFT-D3 treatment of adsorption processes on ionic surfaces. *ChemPhysChem* **2011**, *12*, 3414–3420.

(33) Martinez, L.; Andrade, R.; Birgin, E. G.; Martinez, J. M. PACKMOL: A package for building initial configurations for molecular dynamics simulations. *Journal of computational chemistry* **2009**, *30*, 2157–2164.

(34) Rackers, J. A.; Wang, Z.; Lu, C.; Laury, M. L.; Lagardere, L.; Schnieders, M. J.; Piquemal, J.-P.; Ren, P.; Ponder, J. W. Tinker 8: software tools for molecular design. *Journal of chemical theory and computation* **2018**, *14*, 5273–5289.

(35) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio Jr, R. A.; others Current status of the AMOEBA polarizable force field. *The journal of physical chemistry B* **2010**, *114*, 2549–2564.

(36) 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. *The Journal of Chemical Physics* **2020**, *152*, 104104.

(37) 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.

(38) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. *Journal of computational chemistry* **2011**, *32*, 2319–2327.

(39) Gowers, R. J.; Linke, M.; Barnoud, J.; Reddy, T. J.; Melo, M. N.; Seyler, S. L.; Domanski, J.; Dotson, D. L.; Buchoux, S.; Kenney, I. M.; others MDAnalysis: a Python package for the rapid analysis of molecular dynamics simulations. Proceedings of the 15th python in science conference. 2016; p 105.